2013

DNA barcoding of Jatropha curcas

Terence L. Stephens
Eastern Washington University

DNA Barcoding of *Jatropha curcas*

A Thesis

Presented To

Eastern Washington University

Cheney, Washington

In Partial Fulfillment of the Requirements

for the Degree

Master of Science

By

Terence L. Stephens

Spring 2013

MASTER'S THESIS

In presenting this thesis in partial fulfillment of the requirements for a master's degree at Eastern Washington University, I agree that the JFK Library shall make copies freely available for inspection. I further agree that copying of this project in whole or in part is allowable only for scholarly puposes. It is understood, however, that any copying or publication of this thesis for commercial purposes, or for financial gain, shall not be allowed without my written permission.

Signature______________________________

Date__________________________________

Abstract

DNA barcoding involves sequencing a standard region of DNA to be used as a tool for rapid species identification. The cytochrome C oxidase gene (COXI) is widely accepted as the standard barcode for animals and currently there is no barcode for land plants. The Consortium for the Barcode of Life (CBOL) has proposed two regions of DNA to be used as a dual-locus barcode. In this study I examine the two proposed barcoding regions, ribulose-1,5-bisphosphate carboxylase/oxygenase (*rbcL*) and maturase K (*matK*) for their ease of amplification, discrimatory power, and sequence quality in *Jatropha curcas*, *Jatropha podagrica*, *Latuca sativa*, and *Arabidopsis thaliana*. I found that the *rbcL* gene in *J. curcas* and the 3' end of *L. sativa* met the three criteria for DNA barcoding. The *matK* gene was unamplifiable in all plants used in this study. The *rbcL* gene is a good candidate for use as a DNA barcode based on its discriminatory power and sequence quality while more efforts are needed to find an alternative to the *matK* gene.

I. *Introduction*

DNA Barcoding

The Consortium for the Barcode of Life (CBOL) is an international organization supporting the establishment of a DNA barcode for plants and aims to contribute to the global standard and collaborative movement in DNA barcoding. The DNA barcode uses a select set of gene sequences to assign unknown individuals to a species (Moritz and Cicero 2004). The process of DNA barcoding compares a reference library of genes of a known species to the organism of unknown origin to determine the degree of similarity between the two. A DNA barcode for any species is akin to a barcode assigned to a product in a supermarket. It serves the same function of identifying a specimen. Barcoding is not for phylogenetic resolution but simply to identify unknown species or to reveal inconsistencies between molecular variation and current perceptions of species boundaries (Moritz and Cicero 2004). Botanists are not the only potential users of DNA barcoding, as the technique can be useful to scientists from other fields as well, e.g. forensic science, biotechnology, food industries, and animal science (Valentini 2008). As more and more genome sequences are being added to public databases such as GenBank, use of barcodes can help facilitate the design of universal primers for PCR based gene amplification and gene sequences to identify an unknown organism. Occasionally, databases can contain mistakes and inaccuracies causing distortion of the end results (Demkin 2009). The sequences containing mistakes are often corrected by other scientists examining the same query.

The efforts of CBOL in standardizing a region of DNA for barcoding can have a valuable impact for ecologists since barcoding databases are collecting not only the gene sequences, but also the primer sequences, taxonomic information, place and date of collection, and specimen images. Figure 1 shows the process of DNA barcoding and its uses. Taberlet (2007) describes five criteria for an ideal DNA barcoding system:

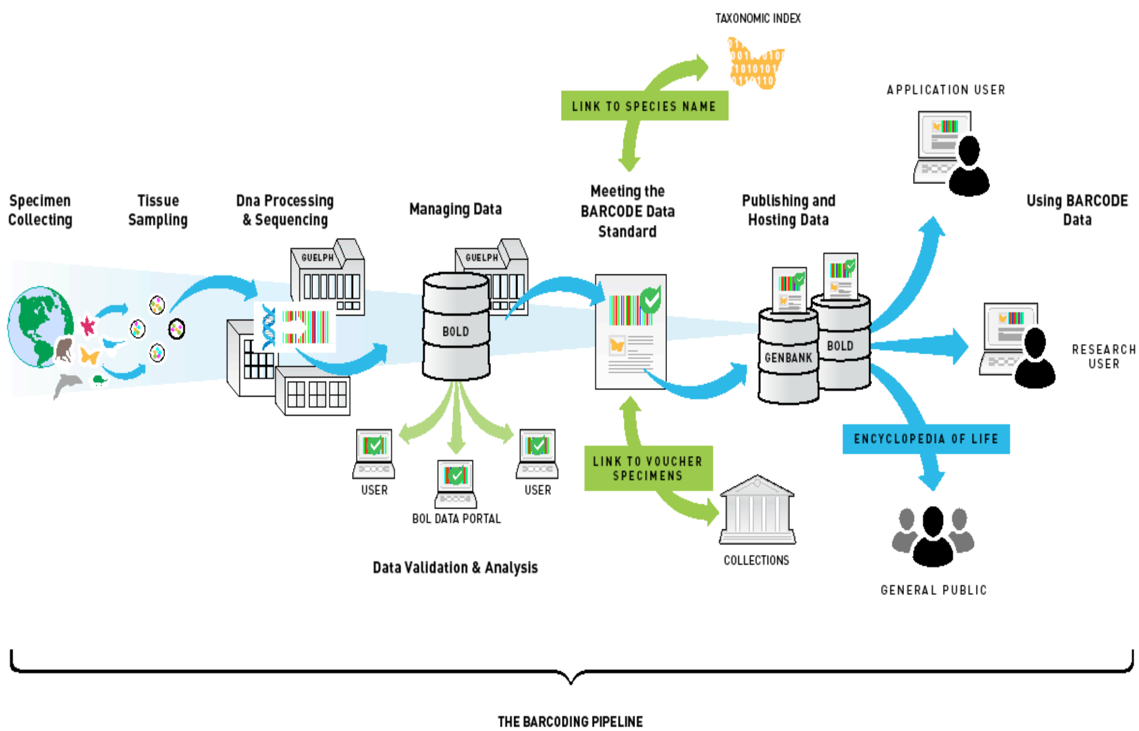


Figure 1. The process of DNA barcoding from initial collection to sequence analysis from http://www.barcodeoflife.org/content/about/what-dna-barcoding.

1. The gene sequences should be nearly identical among individuals of the same species, but different between different species.
2. The proposed barcode should represent select, mutually accepted genes from the same genome.
3. The target DNA region should contain enough phylogenetic information to easily assign unknown or not yet ‘barcoded’ species to their taxonomic group (genus, family, etc.).

4. The proposed barcode should be extremely robust, with highly conserved priming sites for reliable DNA amplifications and sequencing. This is particularly important when using environmental samples, where each extract contains a mixture of many species to be identified at the same time.
5. The target DNA region should be short enough to allow amplification of degraded DNA. Usually, DNA regions longer than 150bp are difficult to amplify from a degraded DNA sample.

For animals, the standard region of DNA for barcoding is a 658 base pair mitochondrial gene, cytochrome *c* oxidase I (COXI). Hebert et al. (2003) found that sequence divergences at the COXI regularly enabled the discrimination of closely allied species in all animal phyla. Except for *Cnidaria*, which has lower COXI divergences than any other phylum used in this study. In another study, Hogg and Hebert (2004) evaluated sequence diversity of the COXI gene to resolve differences among species of *Collembola*. The *Collembola* examined represented all major families found in the Arctic. Members of 14 genera and 19 species were examined and all the organisms were successfully discriminated to the species level with sequence divergences of less than 1% and at the genus level with greater than 8% sequence divergence. This study showed that DNA barcoding can be a useful tool for identifying species of *Collembola* using the mtDNA COXI gene. In most land plants, the mtDNA COXI is not suitable for DNA barcoding because the mitochondrial genes in plants evolve too rapidly to allow for accurate discrimination. Barcoding of nuclear DNA (inherited from both parents) would provide more information than chloroplast DNA (cpDNA), but chloroplast DNA is easier to sequence without cloning and is less likely to occur in multiple copies when compared to the nuclear genome (Chase et al. 2005). Nuclear DNA also exhibits high rates of substitutions, which makes referencing unknown species difficult.

Similar to Taberlet, CBOL selected three criteria for finding a proposed DNA barcode for land plants: *Universality*, i.e. which loci can be easily sequenced from land plants; *Sequence quality and coverage*, i.e. which loci will produce bidirectional sequences with no or few ambiguous base calls; and *Discrimination*, i.e. which loci will allow most species to be distinguished from one another (CBOL Plant Working Group 2009). From a pool of various laboratories, they retrieved sequence data of 907 samples representing 445 angiosperms, 38 gymnosperms, and 67 cryptogam species. Using these samples, they evaluated 7 possible sequence candidates: four are parts of coding genes (*matK, rbcL, rpoB,* and *rpoC1)* and the other three are the noncoding spacer regions (*atpF-atpH, trnh-psbA,* and *psbK*-psbI). From these 7 loci, two short coding regions of cpDNA (*matK* and *rbcL*) have been selected as proposed DNA barcodes by the CBOL Plant Working Group (2009). On average, 72% of plant species can be distinguished using a two-locus barcode. By using the same two locus barcode for known species in the same geographic region, a greater degree of discrimination is possible approaching 100% in some cases (Chase and Fay 2009).

Chloroplast

The well conserved chloroplast genome is the starting point for many research studies involving plants. Chloroplasts are photosynthetic organelles that provide essential energy for plants and algae and are believed to have originated through endosymbiosis between a photosynthetic bacterium and a non-photosynthetic host (Stit 2010). Chloroplasts, also known as plastids, contain hundreds of different proteins that carry out a variety of metabolic functions involving starch metabolism, biosynthesis of amino acids, chlorophyll, fatty acids, etc., in addition to their primary involvement in photosynthesis (Palmer 1987). A single cell in leaves contains anywhere from 400 -1,600 chloroplasts. The angiosperm chloroplast genome varies little in size, structure, and gene content, ranging from 135 to 160 kilobases (kb). Chloroplast genomes contain a large 20 - 30 kb inverted repeat, which divides the remainder of the genome into two regions, one large single copy (LSC) and one small single copy (SSC) region. Most of the genes within the chloroplast genome code for photosynthetic proteins, while the remainder are transfer RNA or ribosomal RNA genes. In the 1980s, an explosion of data resulting from two parallel and complementary developments occurred: (1) the intensive sequencing and characterization of 10 chloroplast genomes (six angiosperms, a bryophyte, and three algae), culminated in the complete sequencing of the three land-plant genomes, and (2) the physical and partial gene mapping of more than 1,000 different chloroplast genomes addressing issues of genome and plant evolution (Palmer 1991). As of 2013, there are approximately 115 species of plants whose chloroplast genome sequences have been completed. Among this group, *Nicotiana tabacum, Latica sativa, Oryza sativa, Spinacia oleracea,* and *Arabidopsis thaliana* have served as model organisms in plant molecular biology.

In 1988, Jeffrey Palmer and his colleagues, while studying chloroplast genomes of different plants, found that swelling or shrinkage of the chloroplast genome was occurring primarily in the non-coding inverted repeat regions. They also found a correlation between the size of the genome and the size of variation in inverted repeats (Palmer 1988). Smaller genomes had fewer numbers of repeats whereas the larger genomes such as Geranium had inverted repeats as large as 76 kb. Changes in the genome complexity of a chloroplast can occur by mutations involving the addition of new sequences and deletion of existing ones, rather than by the gradual drifting of repeated elements that eventually become a single large copy. A vast majority of the length polymorphisms are short, ranging from 1 to 10 base pair sequences in non-coding regions. Large scale changes in inverted repeats are less-frequent and is detectable through restriction-fragment length polymorphisms (RFLPs) and involves the additions or deletions of 50 to 1200 base pairs (Fazekas 2008).

The chloroplast genome is generally smaller than that of mitochondrial or nuclear genomes. Land plant cpDNA evolve more slowly and rarely undergoes internal rearrangement when compared to mitochondrial and nuclear DNA. This is contrary to its relative abundance in relation to size of the mitochondrial and nuclear DNA (Palmer 1991). The conserved genome of cpDNA combined with its smaller size suggests that the chloroplast genome may be evolving under strong constraints to prevent or eliminate any unnecessary changes in sequences. Cullis (2005) suggests that one possibility of this phenomenon is that the plants with smaller chloroplast genomes efficiently reject additional DNA by either not incorporating or by rapidly removing DNA sequences which then migrates to the nucleus as part of an overall strategy to maintain a small genome. In some cases cpDNA can migrate to the nucleus with the help of cpDNA encoded proteins. This size constraint is further exemplified by non-photosynthetic flowering plants whose plastid genomes have eliminated most or all of their photosynthetic genes (Palmer 1991).

Chloroplast DNA is a large component of total cellular DNA and has a conservative rate of nucleotide substitution. These slow rates of molecular evolution are ideal for plant studies at or beyond the family level. The characteristics discussed above make the chloroplast genome a viable platform for phylogenetic related studies. Chloroplast genomes are not the only genome from which we can derive genetic information for analysis. Nuclear and mitochondrial genomes can also be used. Although the latter genomes are not desirable because of higher rates of mutations, internal rearrangement, and larger genome sizes.

Chloroplast vs Mitochondrial & Nuclear Genomes

Mitochondrial and nuclear genomes are larger than chloroplast genomes, which can potentially increase inaccuracy and cost of sequencing. Mitochondrial DNA (mtDNA) also evolves very rapidly in comparison to cpDNA. In 1988, Palmer and colleagues reported that mitochondrial and chloroplast genomes contain two sets of vital genes. The first set of genes encode many proteins crucial to the fundamental bioenergetic processes of the cell, and the second set encodes the components necessary for gene expression. Given the importance of these genes, Palmer entertained the idea that these highly valuable processes would change very slowly over the course of evolution, unlike animal mtDNA which evolves very rapidly. Although sequence comparisons of several plant mitochondrial genes suggest a much lower rate of point mutations in plant mitochondria than animal mitochondria, point mutations in plant mitochondria are still higher when compared with point mutations rates in the chloroplast genome. Thus, the mitochondrial genome is less reliable for intra-specific studies in plants.

Nuclear genomes are closely intertwined with the chloroplast DNA. That is the nucleus controls the synthesis of many proteins found in the chloroplast (Cullis 2009). In a sexually reproducing diploid, one allele at each locus in the nuclear genome comes from the mother and the other allele comes from the father, which is known as biparental inheritance. But not all DNA is inherited equally from both parents. Mitochondria and chloroplast DNA are inherited from one parent and are found outside the cell nucleus. Biparental inheritance can be dismal for barcode analysis as the genetic make-up is inherited from two sources compared to chloroplast genomes which come from a single source. Nuclear genomes also are more difficult to work with because they have high rates of nucleotide substitutions and deletions as well as high rates of gene rearrangement (Zhang 1996).

The completed genome sequences of *Arabidopsis thaliana* and *Oryza sativa* have helped researchers search for evolutionary evidence of the transfer of genes from the original endosymbiont to the nucleus. The proteins encoded by the *Arabidopsis* nuclear genome that are closely related to proteins encoded by the chloroplast genomes (44 plastid genes) of other species as identified by the Arabidopsis Genome Initiative provides evidence for genetic transfer between the chloroplast and the nuclear genomes (Cullis 2009). The genes missing from the *Arabidopsis* chloroplast genome but present in the nucleus are presumed to represent organelle-to-nuclear gene transfers which occurred sometime after the divergence of the organelle-containing lineages (Cullis 2009).

The use of DNA based identification of species has increased in the past decade. DNA barcoding is one of the emergent methods for identification of unknown species and to distinguish between various clones within the same species (CBOL Plant Working Group 2009).

Although the animal kingdom has a standard DNA barcode in place, there is no such standard for land plants.

Maturase K

The first of the two DNA barcoding regions proposed by the CBOL Plant Working Group is the Maturase K *(matK)* gene of the chloroplast genome. The *matK* gene, formerly known as *orfK,* is approximately 1500 base pairs long and is located within the intron of the chloroplast gene *trnK* on the large single-copy section adjacent to the inverted repeat (Hilu and Liang 1997). A homology search of the putative ORF's gene product showed that the amino acid sequence at the carboxyl end is similar to portions of maturase-like polypeptides and might be involved in splicing group II introns (Neuhaus and Link 1987). Open reading frames of 509 codons of *matK* were first characterized in tobacco, *Nicotiana tabacum* L., (Sugita et al. 1985). A slightly longer reading frame of 524 codons was later found in mustard, *Sinapis alba L*., (Neuhaus and Link 1987). The presence of *matK* in the parasitic *Epifagus,* a taxon that has lost approximately 65% of its chloroplast genes, speaks for the functional significance of the *matK* gene (Hilu and Liang 1997). The gene *matK* has also been effective in addressing many systematic questions in various species.

For DNA Barcoding, the *matK* gene has evolved at a higher rate than most of the other genes within cpDNA currently used in plant systematic studies. Olmstead and Palmer (1994) reported that out of 20 genes used in molecular systematics, *matK* had the highest nucleotide substitution rate. Johnson and Soltis (1994) reported that the *matK* gene evolves approximately three fold faster than the *rbcL* gene. Comparing eleven complete sequences representing multiple families and nine partial sequences representing monocot families from GenBank, Hilu and Liang (1997) showed that *matK* had a relatively high rate of substitution in the conserved regions of the gene. They ultimately concluded that the relatively conserved 3' region and the less conserved 5' region provided two sets of characteristics that can be used to discriminate between and within species. In terms of amplification, a sequencing success of 85-88% was found for the *matK* gene region through the use of up to 10 combinations of primers or with more sophisticated chemistry at the amplification stage (Piredda 2011).

Maturase K is the most rapidly evolving plastid coding region and regularly shows high levels of discrimination (CBOL Plant Working Group 2009). In contrast to the reports by others who found that the *matK* gene region was difficult to amplify (Wicke 2009), the CBOL Plant Working Group reported that 90% of the angiosperms tested in 2009 were successfully distinguished from one another using a single primer pair to amplify and sequence the *matK* gene (CBOL Plant Working Group 2009). The transition-transversion ratio for *matK* gene sequences was calculated to be 1.25, which is lower than the expected value of 2.0 for relatively recently diverged sequences and exceeds the value by 0.4 for highly substitution-saturated sequences (Homquist 1983). The higher rate of substitution, reasonable size, and higher level of discrimination underscore the usefulness of the *matK* gene for DNA barcoding purposes.

Ribulose-1,5-bisphosphate carboxylase/oxygenase

The 1400 base pair Ribulose-1,5-bisphosphate carboxylase/oxygenase gene, also known as RubisCO or *rbcL* (*L*=large subunit), is responsible for fixation of carbon dioxide in plants. The enzyme RubisCO consists of eight identical chloroplast-encoded large subunits and eight small subunits (Clegg 1993). RubisCO is the most abundant protein found on earth and can comprise up to 50% of the total soluble protein found in leaf tissue (Tabita 2007). This gene is responsible for the first steps in photosynthesis in plants and algae and has been well conserved throughout evolution. Some chloroplast-encoded genes are interrupted by introns but this is not the case with RubisCO, which illustrates one of several important advantages of the *rbcL* gene (Clegg 1993).

Another advantage of using the *rbcL* gene as a DNA barcode is its ability to be used as a two locus barcode with the gene *matK*. Using *matK* as one of the dual barcodes can allow for better intra-specific species studies because of its higher rate of variability, while *rbcL* will allow for higher discriminatory power between species because of its conserved nature. Direct estimates of nucleotide substitution rates for *rbcL* confirmed this (Zurawski 1984). Because of the conserved rate of nucleotide substitutions, universal primers which should be applicable to nearly all flowering plants can be constructed for use in the polymerase chain reaction (PCR).

Hasebe and colleagues (1994) obtained *rbcL* sequences from the chloroplast genome of Pteridophytes, vascular plants that have the greatest loss of phylogenetic information because of their long evolutionary history. Two phylogenetic trees were developed by the neighbor-joining method and the parsimony method using *rbcL* gene sequences from 58 species representing almost all families of the Leptosporangiate ferns. These two methods produced almost identical phylogenetic trees, which gave three new insights into the evolutionary trends of Leptosporangiates: 1) two morphologically distinct heterosporous water ferns, *Marsilea* and *Malvinia*, are sister genera; 2) the tree ferns (Cyatheaceae, Dicksoniaceae, and Metaxyaceae) are monophyletic; and 3) polyploids are distantly related to the Gleichenioids in spite of the similarity in morphologies. This shows that the high rate of conservation in *rbcL* is important within photosynthetic plants. In addition, sequence variation in the *rbcL* gene is easily amplified by PCR. The CBOL Plant Working Group (2009) found a 90-98% success in PCR and DNA sequencing using a single pair of primers in angiosperms. The *rbcL* gene is capable of providing high levels of discriminatory power, which increases the potential of DNA barcoding.

Throughout the chloroplast genomes, the *rbcL* gene is well characterized. With the improvements in designing universal primers, discriminatory ability of the *rbcL* gene sequence, and bidirectional sequencing for the elimination of ambiguous base calls (CBOL Plant Working Group 2009), *rbcL* is a valued choice in DNA barcoding schematics. Although not the most variable region, it is a frequent component of the best performing multi-locus combinations for species discrimination (CBOL Plant Working Group 2009).

Polymerase Chain Reaction

In 1983, Dr. Kary Mullis successfully invented the process to rapidly amplify a piece of DNA called the polymerase chain reaction (PCR) and as a result was awarded the Nobel Prize in Chemistry for his work. PCR is performed by isolating the DNA from a sample and then using a pair of oligonucleotides, commonly known as primers, to repeatedly amplify target DNA. Primers can be anywhere from 15 - 35 base pairs (bp) long and are a necessary starting point for DNA synthesis.

There are three steps in PCR: denaturation of the DNA, annealing with primers, and the extension of the primers. The first step is performed by increasing the temperature of the DNA sample to 95°C so that the hydrogen bonds in the DNA break and the double stranded DNA becomes a single stranded DNA. The next step involves reducing the temperature between 45 and 60°C to allow the primers to anneal to the complementary strand of single stranded DNA. The primers will attach to the flanking ends of the target gene sequence. The final step uses DNA polymerase and the free nucleotides that have been included in the reaction to build the DNA by extending the primers, generally at 72°C. Since each round generates twice the number of gene copies, the number of copies of the parental gene increases exponentially with every cycle. This process is repeated until an ample amount of DNA is synthesized or obtained, usually after 30-35 cycles. Figure 2 shows the polymerase chain reaction cycle.

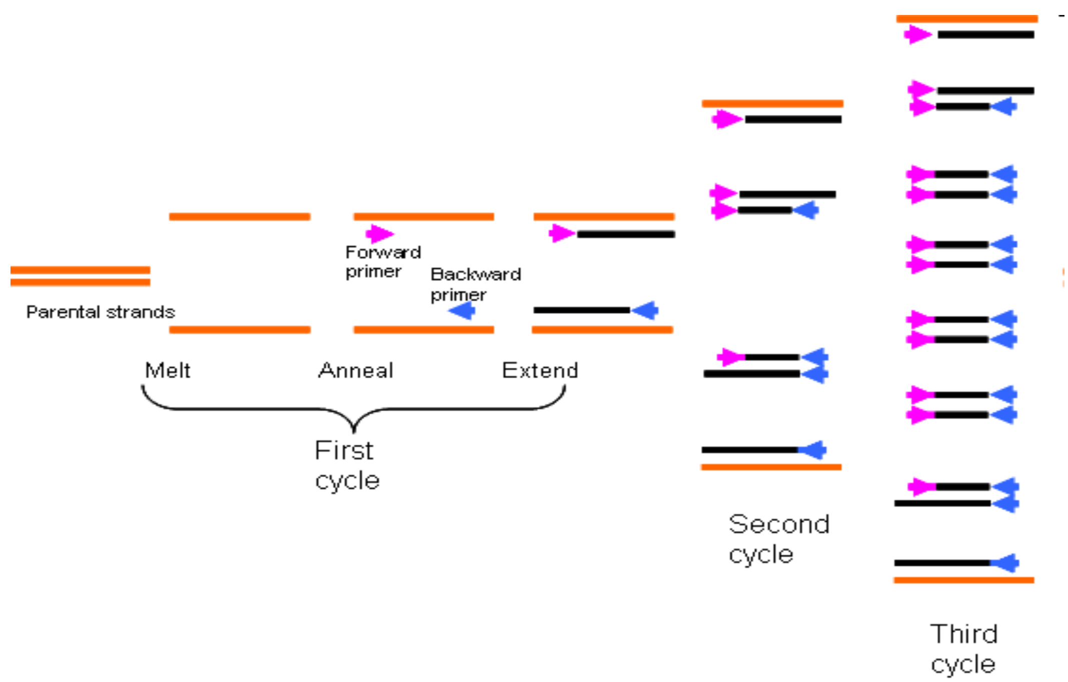


Figure 2. Polymerase chain reaction cycle. The first cycle shows the melting of the two strands creating 2 single strands to which primers anneal to followed by extension of the primer to synthesize the complimentary strand. Black lines are the complimentary strand and the arrows represent primers (http://www.obgynacademy.com/basicsciences/fetology/genetics/*).*

A crucial part of PCR is the choice of the primers to be used. There are two classes of primers, species-specific and universal (Vanhoutte 2004). Species-specific primers amplify target sequences from one species or from possibly a few closely related species and are generally used only if the target sequence is known. Although species-specific primers require more work to design because preliminary information about the gene to be amplified is needed, these primers will decrease the likelihood of amplifying undesirable non-specific DNA (Atkins et al. 2005).

The universal primers are meant to amplify the same region of DNA in a variety of species. This is achieved because homologous sequences in different species often show a degree of similarity to one another via a descended ancestral gene. Examples of this homology can be seen in databanks such as the National Center of Biotechnology Information (NCBI) and the European Molecular Biology laboratory (EMBL). These are large public databases that contain thousands of DNA sequences that represent a wide variety of taxonomic groups used for the classification of many ancestral and recent organisms. Universal primers are ideal for species that lack known sequence data.

Jatropha curcas

Karl von Linne first classified *Jatropha curcas* in 1753 and placed the first existence of *Jatropha* around 70 million years ago (Becker and Makker 2008). This plant is the most primitive member of the Euphorbiaceae family, which contains 2 subgenera, 10 sections, 10 subsections, and up to 175 different species (Becker and Makkar 2008). *Jatropha curcas,* commonly referred to as the physic nut, can grow up to 9 meters tall and is native to tropical America, but is now found in many tropical and sub-tropical regions throughout Africa and Asia.

This dispersal is most likely a result of distribution by Portuguese ships via the Cape Verde islands and Guinea Bissau, where *Jatropha* oil imported from Cape Verde was used for soap production and lamps (Heller 1996). Several factors influencing the distribution of *J. curcas* could be its drought resistance, ease of clonal propagation, short maturation period and wide range of climatic tolerances. In addition to these highly favorable characteristics, *J. curcas* is capable of maintaining and growing in soils that are unsuitable for traditional agriculture such as rocky and sandy soils.

The seed oil of *Jatropha* can be used to make candles and soap, or it can also be used as an alternative to petroleum-based diesel fuels in vehicles. The latter use suggests an economically attractive trait as a fuel extender because of the increased cost, limited availability, increased world-wide consumption, and pollution associated with fossil fuels. The potential use of *Jatropha* oil as a renewable energy source is particularly important for developing countries that are exploring substitutes for fossil fuels and have limited economical resources or lack of access to fossil fuels. In the United States, the implications of manufacturing *Jatropha* as an alternative to diesel may be more practical due to its ability to better counteract greenhouse effects by reducing emissions from motor vehicles. The production of biodiesel from *J. curcas* is not a labor intensive process, does not require any major modifications to a diesel engine, and has a historical precedent as the seed oil of *Jatropha* was used as a diesel fuel substitute during World War II (Kumar 2008).

In 2003, Pramanik et al. looked into the properties of *J. curcas* oil as diesel fuel blends in compression ignition engines. During this study, blends of varying proportions of *J. curcas* oil and diesel were prepared, analyzed, and compared with the performance diesel fuel in an engine.

In addition, they also looked into the effect of temperature on the viscosity of biodiesel. They found that 40-50% (v/v) of *Jatropha* oil can be substituted for diesel with no engine modification and no preheating of the blends. The current production of biodiesel from *J. curcas* requires less land per unit of biofuel than does biodiesel produced from ethanol (Youngs 2012). A report distributed by Goldman Sachs (Barta 2007) indicated that *Jatropha* biodiesel could be produced at a cost of $43 per barrel, making it competitive with petroleum diesel and placing it below the cost of biofuels based on other crops, such as sugarcane ($45 per barrel), corn ($83 per barrel), and soy beans ($122 per barrel) as shown in figure 3 (Cooke 2009). *Jatropha curcas* is an inedible crop, whereas using edible crops for biofuel in developing countries could decrease food supply. More importantly, *J. curcas* is capable of growing in soils that are not typically used for agriculture. As a result, when used as a source of biofuels, *Jatropha* does not compete with edible food crops which require agricultural land and allows cultivators to utilize lands that would otherwise be considered unprofitable. In 2009, King found that China and India had approximately 6 million acres of land planted with *J. curcas* and estimated it would reach over 23 million acres by 2010 (King 2009). Green Acres Costa Rica, an advanced biofuel plantation in Central and North America and leader in the Costa Rican Jatropha Initiative, shows that worldwide there are approximately 12 million acres planted and, if this trend continues, more than 32 million acres will be planted by 2015, generating an estimate $10 billion (Greenacrescostarica.com 2012). Although these estimates vary in different reports, the trend towards using more land for *J. curcas* is evident.

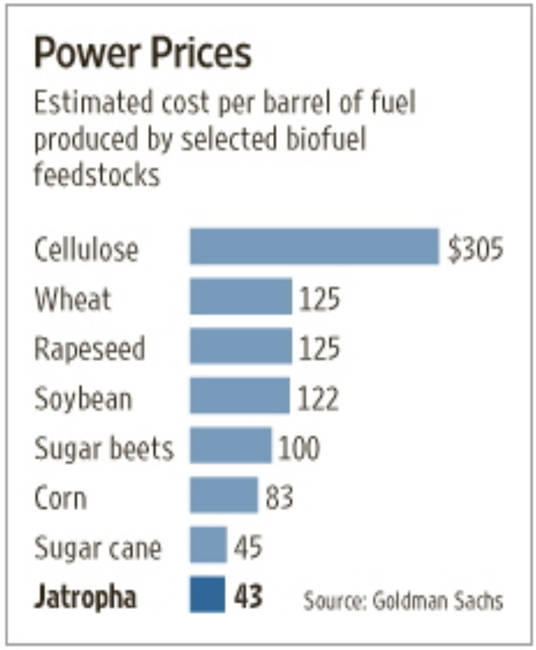


Figure 3. Production cost of various biofuel feedstocks reported by Goldman Sachs (Cooke 2009)

Botanical description of Jatropha curcas

Jatropha curcas is a diploid plant (2n=22) that sheds its leaves during the dry season. This deciduous tree can grow readily from cuttings or seeds, although trees propagated by cutting show a lower longevity and possess a lower drought and disease resistance than those propagated by seeds (Heller 1996). Flowering occurs during the wet season and two flowering peaks are seen during summer and autumn. However, since this plant can maintain and grow in a wide range of climates, flowering occurs throughout the year when *J. curcas* is grown in high precipitation regions, with dormancy induced by fluctuations in rainfall and temperature/light (Heller 1996). The leaves exhibit 5 to 7 lobes with a length of 6 to 15 cm. *Jatropha* is monoecious, the flowers are unisexual, and occasional hermaphrodite flowers can occur (Dehgan and Webster 1979). Pollination of *J. curcas* occurs by insects. It is thought that the heavy perfume, greenish white flowers, versatile anthers, and protruding sexual organs, and copious nectar draw insects to this particular plant (Heller 1996). In a setting such as a greenhouse, the pollination must be done by hand because greenhouse conditions generally do not promote pollination by insects. Seeds are black in color and mature in approximately 3 - 4 months after pollination. Some *Jatropha* plants also contain large amounts of latex. This latex is responsible for the anticoagulant and coagulant properties found in species such as *J. curcas* and *J. podagrica.* For example, investigation of the coagulant activity of the latex from *Jatropha* showed that while whole latex significantly reduced the clotting time of human blood, diluted latex prolonged clotting time (Zhang et al. 2009). In addition, *Jatropha* latex contains a proteolytic enzyme, curcain, which has been demonstrated to have wound-healing properties (Makkar and Becker 2009).

Toxic properties of Jatropha curcas

J. curcas has both toxic and non-toxic varieties. The toxic version is found in South and Central America, Africa, and Asia (Pamidimarri 2009). The recently identified non-toxic version of *J. curcas* is found in Mexico and efforts for cultivating non-toxic versions of *J. curcas* have been underway (Pamidimarri 2009). The main agent responsible for the *Jatropha* toxicity is the high concentration of phorbol esters present in the seed (Makkar 1997). On average, 70% of phorbol esters are retained in the oil and the rest is found in the deoiled kernel meal (Makkar and Becker 2009). Methanol extract of *Jatropha* oil containing phorbol esters has been shown to have strong insecticidal and pesticidal effect against various organisms (Makkar and Becker 2009). When ingested, seeds of the toxic version of *J. curcas* can cause purgative effects (Becker and Makkar 2008). Although humans rarely consume the seeds of *J.* curcas*,* there have been incidences of *J. curcas* nut poisoning in humans after accidental consumption of the seeds, resulting in symptoms such as giddiness, vomiting, diarrhea, and in severe cases, death (Becker and Makkar 1998). Similar purgative effects are seen in animals after consumption, which is one reason why farmers grow *J. curcas* as natural hedges around agricultural land to detour invasive animals from eating their crop.

Other uses of Jatropha curcas

The byproduct of *Jatropha* oil extraction is a nutrient rich seed cake. This seed cake can be valuable as organic manure and would simultaneously serve as a biopesticide/insecticide. Seed cake obtained from the toxic *J. curcas* cannot be used for animal feed (Pamidimarri 2009), but it is possible to reduce the toxicity of the meal by chemical treatments (Martinez-Herrera 2006). Figure 4 shows the use of *J. curcas* at various stages during processing.

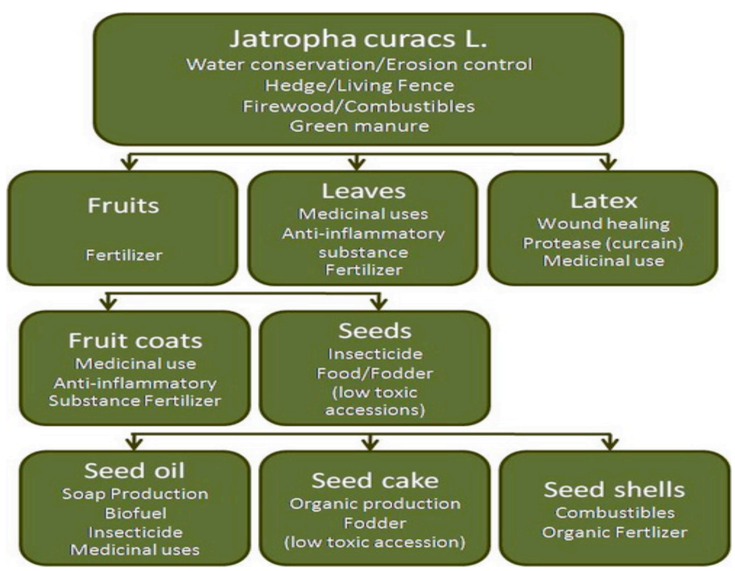


Figure 4. Jatropha curcas - The present uses for the plant (provided by: chemicallygreen.com)

The hardiness and rapid growth of *J. curcas* makes it a prime candidate for erosion control and land reclamation. *Jatropha* also has many medicinal properties, such as treatment of skin diseases, reduction of inflammation, treatment of malaria, use as a contraceptive, healing of wounds, soothing of coughs, and inducement of abortion (Gubitz et al. 1999). However, few of these properties have been carefully studied. *Jatropha curcas* is a multi-purpose plant that has tremendous benefits for the use as an alternative form of biodiesel, as well as serving to provide a wide range of uses at the global and local levels.

With the emergent importance of oil from *Jatropha curcas* as a renewable source of fuel, ways to identify particular traits, family histories, and genetic information are important. Without basic genetic information involving the mechanistic approaches in cultivating this multi-purpose crop, the maximum profit will be difficult to achieve. DNA barcoding provides a simple and fast approach to obtain valuable information needed for the proper enhancement of *J. curcas*. Information such as genetic comparisons, taxonomic classification, morphological differences, family histories, and geographical data are collected and deciphered from the actions of DNA barcoding.

Goals of this study

Two proposed barcoding regions, *rbcL* and *matK*, were used to examine *Jatropha curcas, Jatropha podagrica, Latuca sativa,* and *Arabidopsis thaliana.* The first goal of this study was to design primers for the *rbcL* and *matK* genes based on the *A. thaliana* chloroplast genome (NC_000932.1), and Primer3 software. This was done to introduce internal controls by way of overlapping primer pairs and to use a widely referenced organism in plant studies to determine primer universality. The second goal was to amplify both gene sequences with the designed and published primers recommended by CBOL to compare the product obtained with the primers used by other scientists. The third goal was to amplify the whole gene sequence of *rbcL* using a single set of modified primers as it has been shown that in some cases multiple pairs of primers must be used to amplify whole gene sequences (Kress 2007). The last goal was to use the primers with other plants whose chloroplast genome is sequenced and available on GenBank and compare the results with the known sequences on GenBank. The reason for doing this was to assure that the sequences obtained were of high quality. I hypothesized that both *rbcL* and *matK* gene regions can be amplified by PCR with relative ease. Sequence comparisons of the amplified gene product with those stored in GenBank, would allow me to determine sequence similarities. Sequence matching of 90% or greater of the nucleotide positions will allow for discrimination between two species or 2 genera.

II. *Materials and Methods*

Jatropha curcas seeds were obtained from Dr. Prakash H. Bhuta. A single *J. podagrica* plant was donated by Brian Green from Spokane Community College Spokane, WA. When the plant was received, it was in its juvenile stage approximately one month after germination. *Arabidopsis* seeds were obtained through the EWU Biology Department. *Jatropha curcas, Jatropha podagrica,* and *Arabidopsis thaliana* plants were grown in the greenhouse at Eastern Washington University (EWU) Cheney, WA. The greenhouse temperature averaged between 68°F and 75°F. *Latuca sativa* leaves were collected at a local grocery store one the day prior to PCR amplification. DNA was amplified by PCR using my designed primers and the published primers recommended by CBOL to compare the quality of sequences obtained using the described primers. Henceforth, the primers I designed using Primer3 are referred to as '*designed*' primers and the primers synthesized using published sequences from CBOL are labeled as '*published*' primers. The '*modified*' *rbcL* primers consist of a combination of the forward *published* primer and reverse *designed* primer of primer pair 4.

Primer Design

The *designed* primers used in this study were generated using Primer3 (version 0.4.0) based on the gene sequence of *A. thaliana* obtained from GenBank for both *rbcL* [NC_000932.1 (54958..56397)] and *matK* [NC_00932.1 (2056..3570)] (table 1). The first set of *designed* primers was constructed in 4 different pairs spanning the 1400 bp region of *rbcL.* The second set of *designed* primers consisted of 4 different pairs spanning the 1500 bp region of *matK*. These primers designed from the *Arabidopsis* chloroplast genome sequence were constructed to allow for an overlap of regions at the beginning and end of each primer, which acted as an internal control of DNA sequencing. After sequencing the amplicon obtained with overlapping primers, eg. primer pair 1 and primer pair 2, the 3'- end sequence of primer pair 1 (forward) should be identical to the sequence obtained with primer pair 2 (reverse). The primers were also designed to flank the targeted gene by 50-100 bp on each end to ensure the whole gene was retrieved.

	rbcL *(first set)*	**Product Size**	**matK** *(second set)*	**Product Size**
Primer Pair 1	F: CTTAATTCATGAGTTGTAGGGAGGG R: CATCGGTCCACACAGTTGTC	**246**	F: AGTTATTTAGAGAAAATTTCGATTTCA R: TTTTTGAAAAGATTGGGTTCG	**201**
Primer Pair 2	F: GACAACTGTGTGGACCGATG R: CCACCGCGAAGACATTCATA	**390**	F: TCCGAACCCAATCTTTTCAA R: GATGGGCCTCTTCTGATGAA	**514**
Primer Pair 3	F: CGCCTCATGGTATCCAAGTT R: CGCGGTGAATGTGAAGAAGT	**435**	F: CCCATCTTTTGAAGCCAGAA R: CGATCAATTCATTCAATATTTCC	**491**
Primer Pair 4	F: GGGGATTCACCGCAAATACT R: GATTGGGCCGAGTTTAATTGC	**669**	F: CAATAGCGAAGGGTTTGGAC R: CTGTATCGCACTATGTATCATTCAC	**594**

*Table 1. Primers based on the sequence of Arabidopsis thaliana (NC_000932.1)***.** *Forward primer (F) and reverse primer (R).*

The *published* primers for *rbcL* and *matK* were selected from previous literature (CBOL Plant Working Group 2009; Hilu and Liang 1997) and designed using Primer3. *Published* primer sequences are described in table 2. The *published rbcL* primers targeted a 634 bp region within the *rbcL* gene. The *matK* primers MG1 and MG15 targeted a 2600 bp region which included the *trnK* exons that flank the *matK* gene on each end. The primers matK5 and matK7b targeted a 300 bp region within the *matK* gene. The product obtained from the *published* primers were sequenced and the results compared to the sequence of the amplicons obtained using my *designed* primers.

CBOL	rbcL	Product Size
rbcLaF	F: ATGTCACCACAAACAGAGACTAAAGC	**634**
rbcLajf634R	R: GAAACGGTCTCTCCAACGCAT	**634**

Hilu and Liang	matK	Product Size
matK5	F: CGATCCTTTCATGCATT	**300**
matK 7B	R: GTATTAGGGCATCCCATT	**300**
MG1	F: CTACTGCAGAACTAGTCGGATGGAGTAGAT	**2600**
MG15	R: ATCTGGGTTGCTAACTCAATG	**2600**

Table 2. Previous literature primers (CBOL Plant Working Group 2009; Hilu and Liang 1997).

Control primers accompanying the Phire Plant kit targeted a 300 bp control DNA fragment in plants. This control was used to ensure that the PCR reaction was working properly. The control primers are shown below in table 3.

Phire Plant	Control	Product Size
CP1	F: AGTTCGAGCCTGATTATCCC	**297**
CP2	R: GCATGCCGCCAGCGTTCATC	**297**

Table 3. Control Primers (Phire Plant Direct PCR Kit by Finnzymes)

The *designed* primers were synthesized by Integrated DNA Technologies (Coralville, IA). They were dissolved in sterile milliQ water to reach a final concentration of 100 picomoles per microliter (= 10x stock solution). Thirty microliters from this stock was added to a separate 1.5 ml eppendorf tube, followed by the addition of 270 microliters of milliQ water into the same 1.5 ml eppendorf tube, creating a 1x working solution, making the final concentration to 10 picomoles per microliter, which was stored at 4°C

PCR Kits

PCR using chloroplast DNA was carried out with the Phire Plant Direct PCR kit or Clonetech's Terra Direct PCR kit. The Phire Plant kit by Finnzymes (cat no. F-130) and the Clontech kit (cat no. 639269) are designed to carry out PCR directly from harvested plant leaves without prior DNA purification. The Phire Plant kit was used as the main kit for PCR experiments. The Clontech kit was used to obtained different results. The Phire Plant kit included their proprietary Phire Hot Start DNA Polymerase, 2x plant buffer, dNTP's, sterilized H_20, and 0.5 mm Harris Uni-Core puncher and mat for retrievel of plant tissue. The Clontech kit included Terra DNA polymerase and 2x Terra buffer. The preparations for 20 microliter reactions using the Phire Plant kit was as follows: 10 micrioliters of 2X Phire Plant PCR buffer, 1 microliter each of the forward primer and reverse primer (10 picomoles per microliters), 0.4 microliters of Phire Hot Start Plant DNA polymerase, sufficient distilled water to make a final volume of 20 microliters, and one 0.5 mm punch of plant tissue. The preparation for 20 microliter reactions using the Clontech kit was as follows: 10 microliters of 2x Terra PCR buffer, 1 microliter each of the forward primer and reverse primer (15 picomoles per microliter), 0.4 microliters of Terra DNA polymerase, sufficient distilled water to make a final volume of 20 microliters, and one 0.5 mm diameter punch of plant tissue.

The PCR preparation started by harvesting young bright green leaves 24 hours before the start of experiment to reduce the starch content prior to DNA extraction. All the PCR tubes and enzymes were kept on ice during preparation. A master mix of the polymerase, buffer, and water was created in a sterile 1.5 milliliter eppendorf tube and 18 microliters of this master mix was pipetted into sterilized PCR tubes. One microliter of each primer set (10 picomoles/microliter) was added to its respective PCR tube.

Plant leaves were sterilized by wiping their surface with 70% ethanol (EtOH) saturated paper towel. Once the leaf was dry, a sterile 0.5 mm Harris Uni-Core puncher was used to punch a disc from the leaf. The 0.5mm disc was then inserted into a PCR tube. Between each sample, the puncher was sterilized by rinsing with 70% EtOH. At the end, each PCR tube had 20 microliters of enzymes and a 0.5 mm disc of the plant leaf. The tubes were then were placed in a TC-572 thermocycler and the thermocycling conditions as described below were applied to amplify the target sequence.

Thermocycling Conditions

Thermocycling conditions were performed as recommended by the manufacturers of the Phire Plant and Clontech kit and is shown in table 4 and table 5, respectively. The *published* primers for *rbcL* recommended specific thermocycling conditions, which are described in table 6. For the *matK* locus, a gradient of annealing temperatures had to be used to determine appropriate annealing temperature. These conditions are shown in table 7.

Phire Plant	Cycles	Time	Temp
Initial Denaturation	1	5 min	98°C
Denaturation	30	30 sec	94°C
Annealing	30	45 sec	50°C
Extension	30	45 sec	72°C
Final Extension	1	10 min	72°C
Hold	∞	∞	4°C

Table 4. Thermocycling conditions for Phire Plant Direct PCR Kit (2-step protocol)

Clontech	Cycles	Time	Temp
Initial Denaturation	1	2 min	98°C
Denaturation	35	10 sec	98°C
Annealing	35	15 sec	60°C
Extension	35	45 sec	68°C
Final Extension	1	10 min	68°C
Hold	∞	∞	4°C

Table 5. Thermocycling conditions for Clontech.

Published *rbcL*	Cycles	Time	Temp
Initial Denaturation	1	4 min	98°C
Denaturation	5	30 sec	94°C
Annealing	5	60 sec	55°C
Extension	5	60 sec	72°C
Denaturation	30	10 sec	94°C
Annealing	30	15 sec	54°C
Extension	30	45 sec	72°C
Final Extension	1	10 min	72°C
Hold	∞	∞	10°C

Table 6. Thermocycling conditions for the published rbcL primers recommended by CBOL Plant Working Group 2009.

matK Gradient	Cycles	Time	Temp
Initial Denaturation	1	5 min	98°C
Denaturation	30	30 sec	98°C
Annealing Gradient (16°C variation)	30	45 sec	58°C
Extension	30	45 sec	72°C
Final Extension	1	10 min	72°C
Hold	∞	∞	4°C

Table 7. Thermocycling conditions for matK gradient. (Gradient temperatures in °C: 47.3, 48.1, 49.4, 50.9, 52.0, 53.3, 54.4, 56.5, 57.4, 59.2, 60.7, 60.8).

Gel Electrophoresis

Agarose gels (Fisher Sci cat no. BP160-100) were made with 50 ml of TAE buffer at a final concentration of 1%. The gel electrophoresis apparatus was obtained from Owl Separation Systems, Inc. (Model B1A class2). Three to five microliters of PCR product were mixed with 2 microliters of 5x CYBR-Gold in 6x loading buffer from New England Biolabs; this was placed in each well of the agarose gel under a thin layer of 1X TAE buffer. The standard 100 bp DNA ladder was obtained from New England Biolabs (cat: N0467G). The electrophoresis was carried out at a constant voltage of 81 volts and ran until the fastest moving band of the loading dye was 1 cm away from the opposite end. Gel photos were viewed using a Kodak EDAS 290 UVP camera.

Sequencing

Once the presence of PCR products and their sizes were confirmed by gel electrophoresis, the products were purified to remove dNTPs and excess primers using the MoBio Ultraclean PCR clean-up kit (cat no. D40505). After the products were purified, they were checked once more for purity and size by running them again on a 1% agarose gel as described above. Their concentration was calculated by visual comparison with a 100 bp ladder. These samples were sent to the University of Washington Biochemistry Department in Seattle, Washington for sequencing. Sample preparation for sequencing is described at http://dnaseq.bchem.washington.edu/bdsf/. Sequencing data was sent back to our lab and the data was viewed using BLAST analysis software available on the NCBI website at http://www.ncbi.nlm.nih.gov/. All sequence data that showed unambiguous sequences were then viewed by Finch TV for visual confirmation and then edited, joined, and aligned using CLC Sequence Viewer 6.

Testing Parameters

All sets of *designed* primers and *published* primers for *rbcL* and *matK* were tested against *Jatropha curcas, Jatropha podagrica, Latuca sativa,* and *Arabidopsis thaliana* leaves using the Phire Plant kit. In addition to these primers, *modified* primers were used to retrieve the entire sequence of *rbcL*. The *modified* primers consisted of the forward primer from the published CBOL study (ATGTCACCACAAACAGAGACTAAAGC) and the reverse primer of primer pair 4 (GATTGGGCCGAGTTTAATTGC) from our *designed* primers.

Designed, *published*, and *modified rbcL* primers along with *designed* and *published matK* primers descirbed previously were tested using the Clontech kit only against *L. sativa* and *A. Thaliana*; the reason for this is explained in the discussion section. Furthermore, the primers were tested *in silico* for their ability to amplify the expected size of product using BLAST software.

Blast Analysis

Alignments of the *rbcL* genes were performed using CLC Sequence Viewer 6. For sequence analysis using BLAST, I selected the top 5 sequences resulting from sequence comparisons against the *rbcL* gene sequence of *J. curcas* and other plants used in this study. BLAST analysis provided the percent identity of sequences, which is the percentage of nucleotides in our sample matched with the nucleotide sequence of referenced genes. For this study, a base call refers to the DNA sequencers' ability to correctly identify a nucleotide base at a given location. Base call quality was determined by Finch TV software. BLAST also provides the gaps present in compared sequences. A gap is a space introduced into an alignment to compensate for insertions and deletions in one sequence relative to another (NCBI). Included in the analysis is how much of the experimental gene covers the referenced gene, also known as overall coverage. BLAST analysis also provides the expected value or E-value. The E-value is a parameter that describes the frequency of sequence similarity one can "expect" to see by chance when searching a database of a particular size (NCBI). The closer the E-value is to zero, the higher the significance of the match. Restriction sites of whole gene sequences and complete aligned sequences are found in the appendix.

III. <u>*Results*</u>

Rate of successful amplification was determined by the number of successful amplifications per number of attempts. The rate of amplification using the Phire Plant kit shown in table 8.

Origin	Primer	Jatropha curcas	% success	Jatropha podagrica	% success	Lactuca sativa	% success	Arabidopsis thaliana	% success
rbcL									
Designed Primers	Primer Pair 1	2/11	**18%**	0/8	**0%**	1/12	**8%**	0/8	**0%**
	Primer Pair 2	4/9	**44%**	0/8	**0%**	1/16	**6%**	1/7	**14%**
	Primer Pair 3	2/19	**11%**	0/8	**0%**	3/13	**23%**	2/6	**33%**
	Primer Pair 4	4/25	**16%**	1/9	**11%**	0/11	**0%**	0/6	**0%**
Published Primers	CBOL	1/1	**100%**	0/2	**0%**	0/1	**0%**	0/4	**0%**
Modified Primers	Published forward/ Designed reverse	1/1	**100%**	1/2	**50%**	0/2	**0%**	0/2	**0%**
matK									
Designed Primers	Primer Pair 1	0/4	**0%**	0/1	**0%**	0/1	**0%**	0/2	**0%**
	Primer Pair 2	0/1	**0%**	0/1	**0%**	0/1	**0%**	0/2	**0%**
	Primer Pair 3	0/4	**0%**	0/1	**0%**	0/1	**0%**	0/2	**0%**
	Primer Pair 4	0/1	**0%**	0/1	**0%**	0/1	**0%**	0/2	**0%**
Published Primers	MG1 MG15	0/3	**0%**	0/7	**0%**	0/3	**0%**	0/4	**0%**
	Matk5 Matk7b	0/3	**0%**	0/4	**0%**	0/3	**0%**	0/4	**0%**

Table 8. Number of successful amplifications using the Phire Plant kit.

The second group of experiments were conducted using the Clontech kit to retrieve different results because during the latter portion of this research the Phire Plant kit was not producing amplicons of the *rbcL* and *matK* gene. *Jatropha curcas* was not used with the Clontech kit because sequences were already obtained. *Jatropha podagrica* was not used for reasons described in the discussion section. The results are shown below in table 9.

Origin	Primer	Lactuca sativa	% success	Arabidopsis thaliana	% success
rbcL					
Design Primers	Primer Pair 1	n/a	**n/a**	n/a	**n/a**
	Primer Pair 2	0/1	**0%**	0/1	**0%**
	Primer Pair 3	0/1	**0%**	0/1	**0%**
	Primer Pair 4	0/1	**0%**	0/1	**0%**
Publish Primers	CBOL	0/1	**0%**	0/1	**0%**
Design/ Publish	Modified	0/1	**0%**	0/1	**0%**
matK					
Design Primers	Primer Pair 1	n/a	**n/a**	0/1	**0%**
	Primer Pair 2	0/1	**0%**	0/1	**0%**
	Primer Pair 3	0/1	**0%**	0/1	**0%**
	Primer Pair 4	n/a	**n/a**	0/1	**0%**
Publish Primers	MG1 MG15	0/1	**0%**	0/1	**0%**
	Matk5 Matk7b	0/1	**0%**	0/1	**0%**

Table 9. Number of successful amplification using the Clontech kit.

Designed rbcL primers against Jatropha curcas using the Phire Plant kit

Using the *designed* primers for *rbcL*, a total of 12/64 (19%) PCR attempts were successful in amplifying the target sequence in *J. curcas*. Primer pair 2 showed the highest rate of success (44%) and primer pair 3 showed the lowest success rate (11%) (table 8). Gel electrophoresis of the amplified product showed bright bands of the expected product. A higher percentage of amplification results were obtained with *designed* primers with *J. curcas* than any other organism in this study. Figure 5 below shows the primer pairs 1-4 successfully amplifying the *rbcL* gene. The size of an amplicon and correspnding primers used are given in the figure legend.

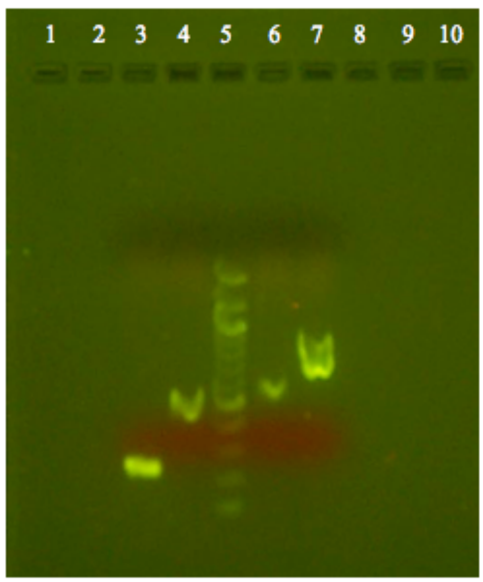


Figure 5. Designed rbcL primers against Jatropha curcas. From well 3 to well 7: primer pair 1 at 240 bp (lane 3), primer pair 2 at 450bp (lane 4), 100bp ladder (lane 5), primer pair 3 at 580bp (lane 6), and primer pair 4 at 690bp (lane 7).

Published and modified rbcL primers against Jatropha curcas using Phire Plant kit

The *published* primers of CBOL and the *modified* primers successfully amplified the *rbcL* gene in a single attempt: a 634 bp region for the *published* primers and a 1400 bp gene region for the *modified* primers (table 8). Figure 6 shows the successful amplification of *rbcL* using the primers from CBOL and the *modified* primers with an expected size of amplicons.

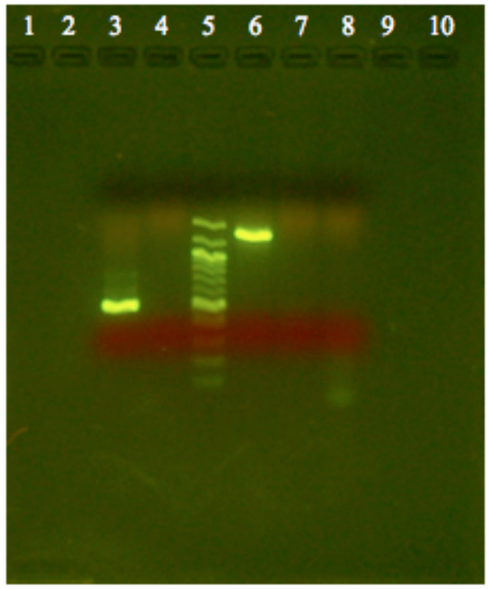


Figure 6. Successful amplification of published and modified primers against J. curcas. Lanes 3 and 4 targeted the 634 bp region (published) while 6 and 7 targeted the whole 1400 bp gene (modified). Lane 8 is the postive control and lane 5 is the 100 bp ladder. Lanes 4, 7, and 8 are without and sample.

Designed, published, and modified rbcL primers against L. sativa using Phire Plant kit

The *designed* primer pair 3 and 4 were successfully used to amplify the *rbcL* gene in *L. sativa*. The 5' end of the *rbcL* gene was less consistently amplified with primer pairs 1 and 2 of *L. sativa.* The amplicons for primer pairs 1 and 2 were visible on the gel but their concentration was too low to sequence. No amplification was achieved using the *published* or *modified* primers against *L. sativa*. Figures 7a, 7b, and 7c show the successful amplification of the second half of the *L. sativa* sequence.

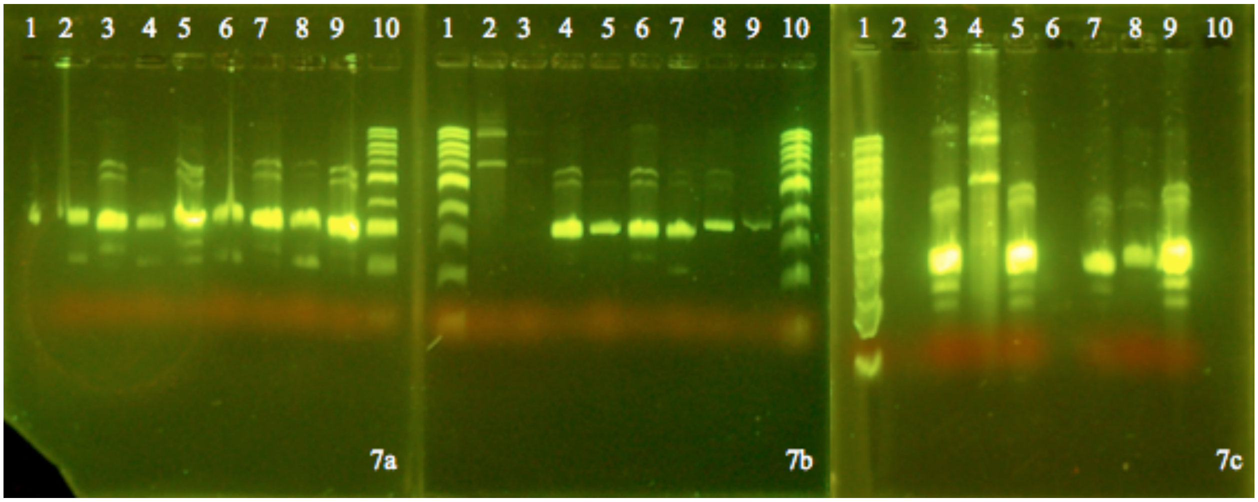


Figure 7a) Lanes 1 through 9 are targeting primer pair 3 (435 bp). 100 bp ladder is shown in lane 10. Figure 7b) Lane 2 and 3 are targeting the primer pair 4 (669 bp) and lanes 4 - 9 are targeting primer pair 3. 100 bp ladders in lanes 1 and 10. Figure 7c) Lane 1 is a 100 bp ladder. Landes 3 - 9 are targeting primer pair 4.

Designed, published, and modified rbcL primers against J. podagrica and A. thaliana using the Phire Plant kit

The *designed* primers for *rbcL* failed to produce any detectable level of amplicon with *J. podagrica* (32 attempts) and with *A. thaliana* (27 attempts). However, of the 27 attempts against *A. thaliana,* three attempts produced very little amplicon giving faint bands on agarose gel, albeit the product was of the correct size. Only *designed* primer pair 2 and pair 3 successfully amplified a portion of the *rbcL* gene of *A. thaliana*. Although some amplification was achieved, the low DNA concentration of *A. thaliana* could not be used for sequencing. The positive control for gene amplifcation included with the PCR kit always showed amplification. Figure 8 shows the failure to see any amplification of *rbcL* in *J. podagrica.*

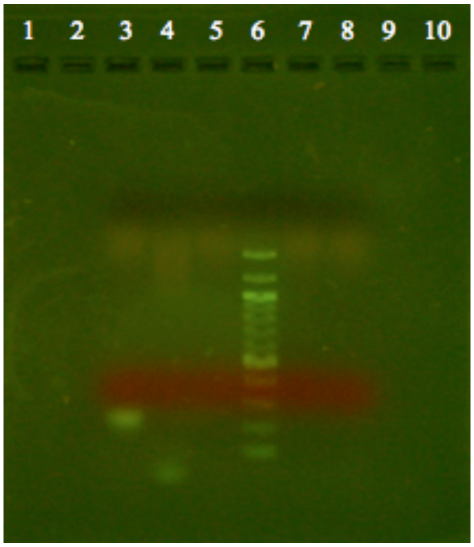


Figure 8. Unsuccessful amplification of J. podagrica using the designed rbcL primers. Lane 4 and 5 targeted primer pair 4 and 3, repsectively. Lane 7 and 8 targeted primer pair 2 and 1, respectively. The 100bp ladder is shown in lane 6 and positive control shown in lane 3.

Amplification of matK gene using the Phire Plant kit

Neither the *designed* nor the *published* primers were able to amplify the *matK* gene from any of the plants tested in PCR. However, the positive control primers once again gave the correct size amplicon. I conducted 16 attempts with *J. curcas,* 15 attempts with *J. podagrica,* 10 attempts with *L. sativa,* and 16 attempts with *A. thaliana* to obtain amplification of the *matK* gene. Figure 9 shows one example of the unsuccessful amplification of *matK* in *J. curcas* using *designed* and *published* primers.

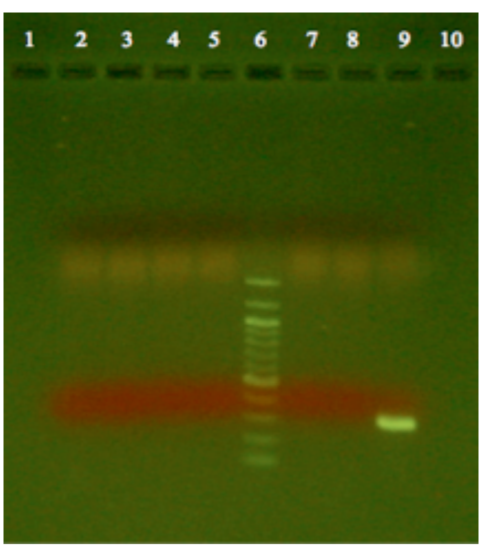


Figure 9. Unsuccessful amplification of matK locus of J. curcas. Lanes 2 through 5 represents use of designed primers and lanes 7 and 8 show results with primers published by Hilu and Liang 1997. A 100 bp ladder is in lane 6 and the positive control is in lane 9.

Unsuccessful amplification using the Clontech kit

No amplification of the *matK* gene was seen with any of our *designed*, *published*, or *modified* primers with *L. sativa* or *A. thaliana* leaves and the Clontech kit. I used *L. sativa* and *A. thaliana* only because in earlier experiments, *J. curcas* was successfully amplified and *J. podagrica* showed difficulty of obtainting amplification.

BLAST results for Jatropha curcas using designed rbcL primers

The *rbcL* gene of *J. curcas* was amplified using the *designed* primers. Gene sequences obtained from their amplicons were joined to create a single strand representing a forward strand and a reverse strand. The strand sequence was used for BLAST analysis to find similar sequences in GenBank and to determine the quality of the *J. curcas* gene sequence. Visual inspection of the sequence showed conserved regions in both strands using the *designed* primers (figure 10). The top 5 hits, or similar sequences, of the complete forward strand of the *rbcL* gene using the *designed* primers revealed a high sequence quality (averages: 97.4% identity, 67.6% coverage) (figure 11).

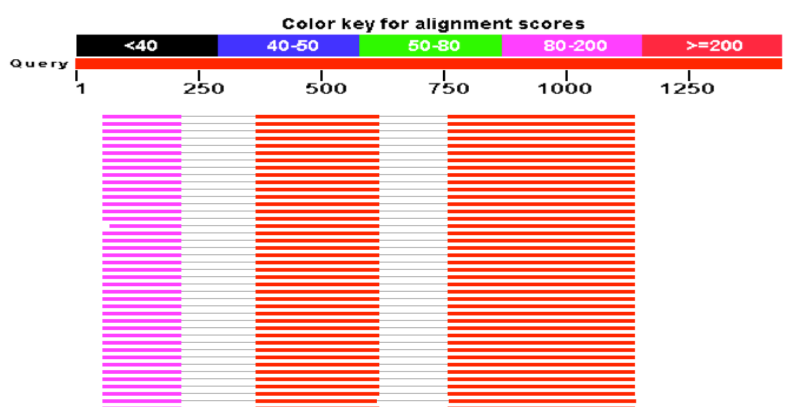


Figure 10. Screen shot of BLAST results showing the conserved regions in rbcL using the designed primers against J. curcas. The color key shows the score of the sequence against the query. The top 5 lines represent the 5 most similar sequences listed in figure 10.

Sequences producing significant alignments:

Select: All None Selected:0

Alignments Download GenBank Graphics Distance tree of results

Description	Max score	Total score	Query cover	E value	Max ident	Accession
Jatropha curcas acetyl-CoA carboxylase beta-CT subunit gene, complete cds	668	1461	68%	0.0	97%	HQ153096.1
Jatropha curcas chloroplast, complete genome	668	1461	68%	0.0	97%	FJ695500.1
Jatropha capensis chloroplast partial rbcL gene for ribulose bisphosphate carboxylase large subu	663	1447	67%	0.0	98%	AM234978.1
Jatropha integerrima chloroplast rbcL gene for ribulose-1,5-bisphosphate carboxylase/oxygenase	663	1450	67%	0.0	97%	AB233879.1
Jatropha integerrima ribulose-1,5-bisphosphate carboxylase/oxygenase large subunit (rbcL) gene	663	1445	67%	0.0	98%	AY794902.1

Figure 11. Screen shot of BLAST analysis using complete forward strand of rbcL using designed primers against Jatropha curcas.

The top five hits of the reverse strand using the *designed* primers revealed a high sequence quality (averages: 99% identity, 55% coverage) (figure 12). An average of 2 gaps was found for both strands.

Sequences producing significant alignments:

Select: All None Selected:0

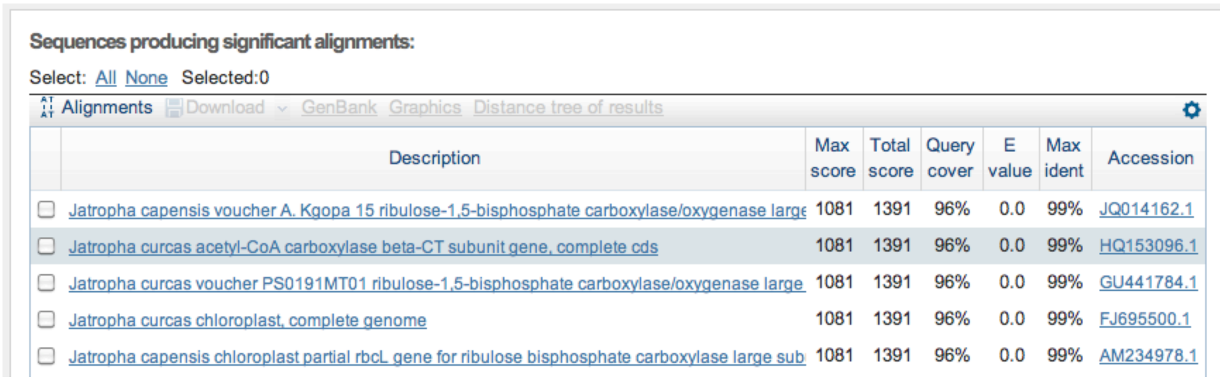
Alignments Download GenBank Graphics Distance tree of results

Description	Max score	Total score	Query cover	E value	Max ident	Accession
Jatropha curcas acetyl-CoA carboxylase beta-CT subunit gene, complete cds	675	1260	55%	0.0	99%	HQ153096.1
Jatropha curcas chloroplast, complete genome	675	1260	55%	0.0	99%	FJ695500.1
Jatropha capensis chloroplast partial rbcL gene for ribulose bisphosphate carboxylase large subu	675	1260	55%	0.0	99%	AM234978.1
Jatropha integerrima chloroplast rbcL gene for ribulose-1,5-bisphosphate carboxylase/oxygenase	670	1255	55%	0.0	99%	AB233879.1
Jatropha integerrima ribulose-1,5-bisphosphate carboxylase/oxygenase large subunit (rbcL) gene	670	1255	55%	0.0	99%	AY794902.1

Figure 12. Screen shot of BLAST analysis using complete reverse strand of rbcL using designed primers against Jatropha curcas.

BLAST results for Jatropha curcas using published rbcL primers

The BLAST result of the forward gene retrieved using the *published* primers revealed a high similarity to plants within the *Jatropha* genus. The forward strand using the *published* primer for the *rbcL* gene had a high sequence quality (averages: 99% identity, 96% coverage) (figure 13). The BLAST results of the reverse strand using the *published* primer showed a high sequence quality as well (averages: 99% identity, 95.2% coverage) (figure 14). An average of 2 gaps was found for both strands.



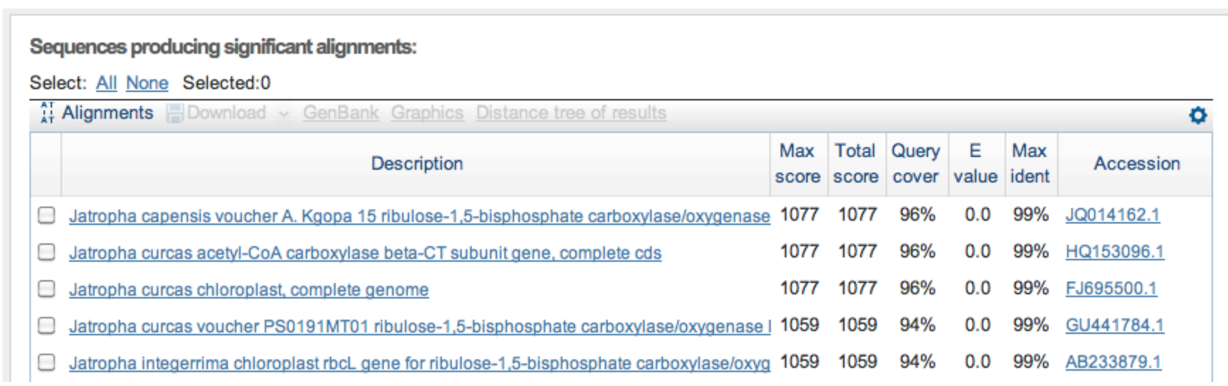
Sequences producing significant alignments:

Select: All None Selected:0

Alignments Download GenBank Graphics Distance tree of results

	Description	Max score	Total score	Query cover	E value	Max ident	Accession
☐	Jatropha capensis voucher A. Kgopa 15 ribulose-1,5-bisphosphate carboxylase/oxygenase large	1081	1391	96%	0.0	99%	JQ014162.1
☐	Jatropha curcas acetyl-CoA carboxylase beta-CT subunit gene, complete cds	1081	1391	96%	0.0	99%	HQ153096.1
☐	Jatropha curcas voucher PS0191MT01 ribulose-1,5-bisphosphate carboxylase/oxygenase large	1081	1391	96%	0.0	99%	GU441784.1
☐	Jatropha curcas chloroplast, complete genome	1081	1391	96%	0.0	99%	FJ695500.1
☐	Jatropha capensis chloroplast partial rbcL gene for ribulose bisphosphate carboxylase large sub	1081	1391	96%	0.0	99%	AM234978.1

Figure 13. Screen shot of BLAST analysis using the published forward primer of the rbcL gene sequence against Jatropha curcas.



Sequences producing significant alignments:

Select: All None Selected:0

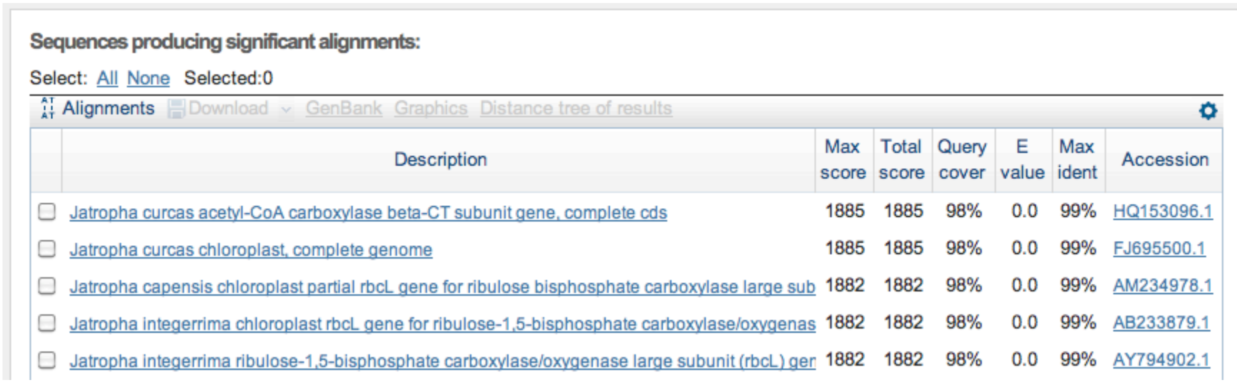
Alignments Download GenBank Graphics Distance tree of results

	Description	Max score	Total score	Query cover	E value	Max ident	Accession
☐	Jatropha capensis voucher A. Kgopa 15 ribulose-1,5-bisphosphate carboxylase/oxygenase	1077	1077	96%	0.0	99%	JQ014162.1
☐	Jatropha curcas acetyl-CoA carboxylase beta-CT subunit gene, complete cds	1077	1077	96%	0.0	99%	HQ153096.1
☐	Jatropha curcas chloroplast, complete genome	1077	1077	96%	0.0	99%	FJ695500.1
☐	Jatropha curcas voucher PS0191MT01 ribulose-1,5-bisphosphate carboxylase/oxygenase l	1059	1059	94%	0.0	99%	GU441784.1
☐	Jatropha integerrima chloroplast rbcL gene for ribulose-1,5-bisphosphate carboxylase/oxyg	1059	1059	94%	0.0	99%	AB233879.1

Figure 14. Screen shot of BLAST analysis using the published reverse primer of the rbcL gene sequence against Jatropha curcas.

BLAST analysis of Jatropha curcas using modified rbcL primers

The primers used in this experiment consisted of the *published* forward primer rbcLaF (ATGTCACCACAAACAGAGACTAAAGC) and primer pair 4 of the *designed* primer (GATTGGGCCGAGTTTAATTGC). The forward strand showed high sequence quality (averages: 99% identity, 98% coverage, and 1 gap) (figure15). The reverse strand gave a high sequence quality (averages: 98.8%, 97.8% coverage, and 8.2 gaps) (figure 16).



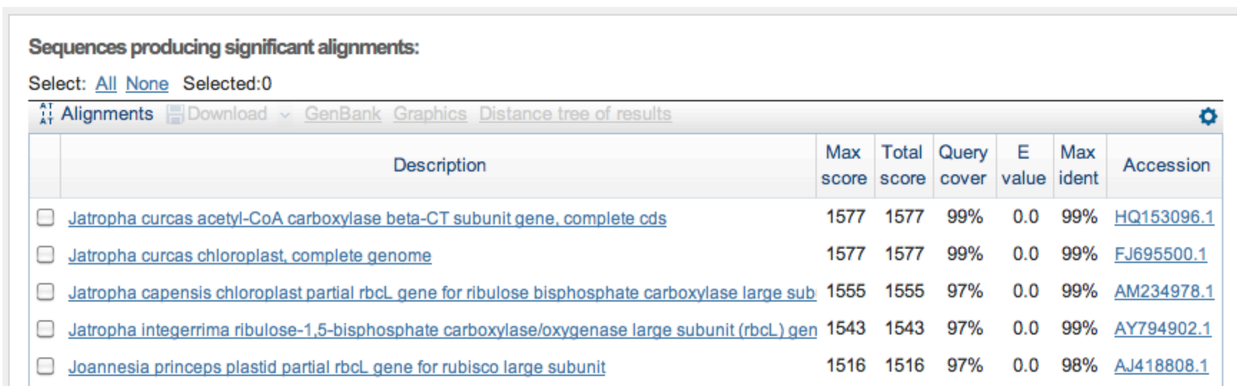
Sequences producing significant alignments:

Select: All None Selected:0

Alignments Download GenBank Graphics Distance tree of results

	Description	Max score	Total score	Query cover	E value	Max ident	Accession
☐	Jatropha curcas acetyl-CoA carboxylase beta-CT subunit gene, complete cds	1885	1885	98%	0.0	99%	HQ153096.1
☐	Jatropha curcas chloroplast, complete genome	1885	1885	98%	0.0	99%	FJ695500.1
☐	Jatropha capensis chloroplast partial rbcL gene for ribulose bisphosphate carboxylase large sub	1882	1882	98%	0.0	99%	AM234978.1
☐	Jatropha integerrima chloroplast rbcL gene for ribulose-1,5-bisphosphate carboxylase/oxygenas	1882	1882	98%	0.0	99%	AB233879.1
☐	Jatropha integerrima ribulose-1,5-bisphosphate carboxylase/oxygenase large subunit (rbcL) ger	1882	1882	98%	0.0	99%	AY794902.1

Figure 15. Screen shot of BLAST analysis using the modified forward primer of rbcL against Jatropha curcas.



Sequences producing significant alignments:

Select: All None Selected:0

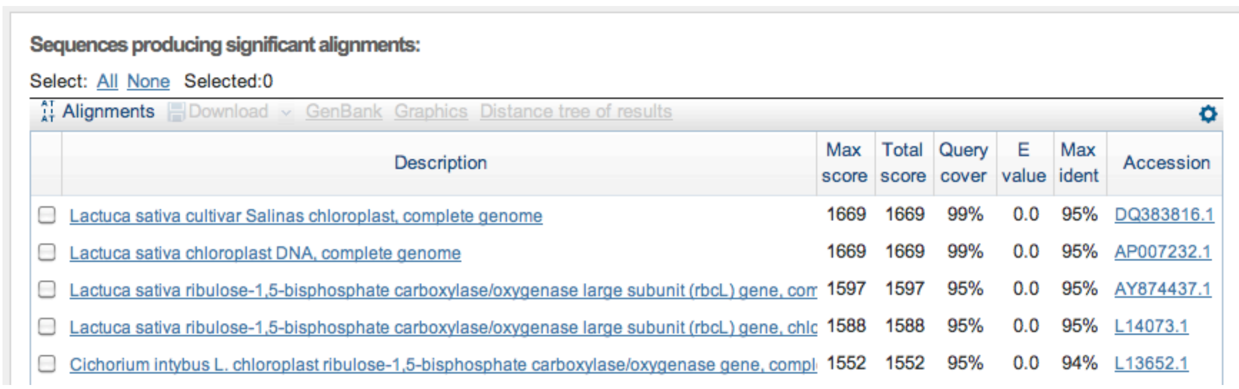
Alignments Download GenBank Graphics Distance tree of results

	Description	Max score	Total score	Query cover	E value	Max ident	Accession
☐	Jatropha curcas acetyl-CoA carboxylase beta-CT subunit gene, complete cds	1577	1577	99%	0.0	99%	HQ153096.1
☐	Jatropha curcas chloroplast, complete genome	1577	1577	99%	0.0	99%	FJ695500.1
☐	Jatropha capensis chloroplast partial rbcL gene for ribulose bisphosphate carboxylase large sub	1555	1555	97%	0.0	99%	AM234978.1
☐	Jatropha integerrima ribulose-1,5-bisphosphate carboxylase/oxygenase large subunit (rbcL) gen	1543	1543	97%	0.0	99%	AY794902.1
☐	Joannesia princeps plastid partial rbcL gene for rubisco large subunit	1516	1516	97%	0.0	98%	AJ418808.1

Figure 16. Screen shot of BLAST anaylsis using the modified reverse primer of rbcL against Jatropha curcas.

BLAST anlaysis of Latuca sativa using designed rbcL primers

Only the second half (3' end) of the *rbcL* gene sequence in *L. sativa* was amplifiable using *designed* primers, therefore only the 3rd and 4th primers were joined together after sequencing to create a 3' end portion. The forward strand of this gene revealed high sequence similarity to closely related species but an increased number of gaps within the alignment (averages: 94.8% identity, 96.6% coverage, and 48.6 gaps) (figure 17). The reverse strand of the joined 3rd and 4th *rbcL* gene region of *L. sativa* revealed similar sequence quality to the *published* forward primer in *J. curcas* (averages: 98.8% identity, 97% coverage, and 2 gaps) (figure 18).



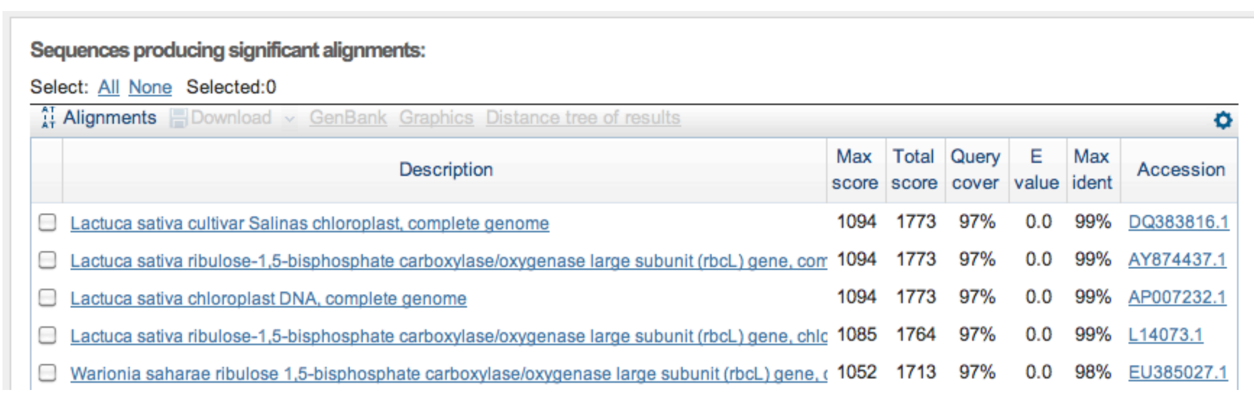
Sequences producing significant alignments:

Select: All None Selected:0

Alignments Download GenBank Graphics Distance tree of results

Description	Max score	Total score	Query cover	E value	Max ident	Accession
Lactuca sativa cultivar Salinas chloroplast, complete genome	1669	1669	99%	0.0	95%	DQ383816.1
Lactuca sativa chloroplast DNA, complete genome	1669	1669	99%	0.0	95%	AP007232.1
Lactuca sativa ribulose-1,5-bisphosphate carboxylase/oxygenase large subunit (rbcL) gene, com	1597	1597	95%	0.0	95%	AY874437.1
Lactuca sativa ribulose-1,5-bisphosphate carboxylase/oxygenase large subunit (rbcL) gene, chlc	1588	1588	95%	0.0	95%	L14073.1
Cichorium intybus L. chloroplast ribulose-1,5-bisphosphate carboxylase/oxygenase gene, compl	1552	1552	95%	0.0	94%	L13652.1

Figure 17. Screen shot of BLAST anaylsis using designed forward primers for the 3' end of L. sativa.



Sequences producing significant alignments:

Select: All None Selected:0

Alignments Download GenBank Graphics Distance tree of results

Description	Max score	Total score	Query cover	E value	Max ident	Accession
Lactuca sativa cultivar Salinas chloroplast, complete genome	1094	1773	97%	0.0	99%	DQ383816.1
Lactuca sativa ribulose-1,5-bisphosphate carboxylase/oxygenase large subunit (rbcL) gene, com	1094	1773	97%	0.0	99%	AY874437.1
Lactuca sativa chloroplast DNA, complete genome	1094	1773	97%	0.0	99%	AP007232.1
Lactuca sativa ribulose-1,5-bisphosphate carboxylase/oxygenase large subunit (rbcL) gene, chlc	1085	1764	97%	0.0	99%	L14073.1
Warionia saharae ribulose 1,5-bisphosphate carboxylase/oxygenase large subunit (rbcL) gene, c	1052	1713	97%	0.0	98%	EU385027.1

Figure 18. Screen shot of BLAST analysis using designed reverse primers for the 3' end of L. sativa.

IV. <u>*Discussion*</u>

DNA barcoding is an effective way to quickly identify unknown species or a member of the same genus as shown by my results (figure 10-16). The CBOL Plant Working Group has shown that the combination of *rbcL* and *matK* gene regions are sufficient to identify species (CBOL Plant Working group 2009). Amplification of these gene regions were accomplished with little to no modification of their primers. The multi-purpose crop *J. curcas* has a wide range of characteristics of which the production of biofuels from the seed oil it produces holds the most practical value. The question of potential barcodes for land plants in combination with the features seen in *Jatropha curcas* has prompted an examination of the proposed barcodes as well as obtaining any additional information for *J. curcas*. My work with *J. curcas, J. podagrica, L. sativa, and A. thalaiana* in the DNA barcoding process is a way to support or oppose the two-locus barcode set forth by CBOL and to characterize these organisms by fixed gene sequences. I found that *designed*, *published*, and *modified* primers for *J. curcas* were successful in amplifying the *rbcL* gene for sequencing and sequence comparison. Multiple attempts of amplifying the *matK* gene were unsuccessful. Various attempts to obtain amplification involved finding an appropriate reannealing temperature, changing the reagent kit used, and using different sized leaves as a source of chloroplast DNA. Under all conditions tested the internal DNA control supplied with the kit amplified well, indicating that the failures to amplify *matK* were not due to malfunctions of the PCR kit. Although *designed* primers for *rbcL* used against the first half of the *L. sativa* gene sequence and *A. thaliana* showed low levels of amplification, the DNA concentrations were too low and prevented further analysis. Similarly, using our *modified* primers against *J. podagrica*, only one attempt out of 37 successfully amplified the gene, but again, the DNA concentrations were too low for sequencing the amplicon.

Analysis of designed rbcL primers for Jatropha curcas

Among the first of four sequences obtained after amplification were used for sequence comparison against *rbcL* genes from GenBank. The 4 primer pairs of the *designed* primers in *J. curcas* required small adjustments in PCR conditions. Using a fresh working stock of primers greatly improved the quantity of PCR product. This was done because the first PCR clean up and sequencing for the amplicon obtained with primer pair 3 gave unclear base calls; the chromatograph of the sequencing results showed multiple peaks or plateaus at the beginning of the sequence instead of a single peak indicating a single nucleotide for a given location. Preparing a new working stock of primers can reduce inadequate priming during the annealing step in PCR, reducing self-priming and ensuring better amplification. Once new primers were made, the second PCR showed a distinct peak for each nucleotide position indicating high base call quality of the sequencing data. The entire 1443 bp gene sequence of *J. curcas rbcL* was aligned and joined using CLC Sequence Viewer 6. The sequenced gene consisted of primer pairs 1, 2, and 4 from the first working stock of primers and primer pair 3 from the second working stock of primers. Primer pair 3 from the second working stock was chosen because of better base call quality. After combining sequences, BLAST analysis was performed and it was found that the top 5 plant genes with 99% sequence similarity were of the closely-related species of *Jatropha*. This is expected since RubisCO evolves at a slower rate and should yield sequence comparison results more similar to that of *J. curcas*. In addition to the sequence similarity, the *designed* primers of the *rbcL* gene of *J. curcas* showed large conserved regions. The implementation of overlapping primers and the clean up of sequenced fragments could account for the large conserved regions as the trimming of ambiguous base calls will not show up in the BLAST alignment. Although only 67% of the reference gene was covered during BLAST, 99% of the base calls were present within the BLAST frame of 1443 base pairs. The first hit acetyl-CoA carboxylase (figure 11) revealed the highest similarity whereas we would expect to see the *rbcL* gene of *J. curcas*. Currently there is no independent entry for the *rbcL* gene of *J. curcas* which is why acetyl-CoA carboxylase is seen first. Additionally, the *rbcL* gene of the whole chloroplast genome of *J. curcas* (figure 11; hit 2) was found within the acetyl-CoA carboxylase reference hit (data not shown) revealing the importance of acetyl-CoA in carbon fixation. The lack of distantly related species in the top 100 sequence (data not shown) hits suggests that the amount and quality of base calls were reliable enough to distinguish *J. curcas* from other plants in GenBank. Appendix 1 shows the fully aligned sequences and restriction sites for both forward and reverse single stranded sequences.

Analysis of published rbcL primers against Jatropha curcas

The *published* primers used against *J. curcas* were successfully and easily amplified. This could be a result of better primer design by CBOL using high quality reference sequences, differences in reagents used, or better pipetting technique. The primers were designed to give a 634 base pair gene region. The forward primer gave 1003 bases and the reverse primer gave 630 bases in the alignment. Both sequences were subjected to BLAST analysis individually and provided high sequence identity. Similar to the *designed* primers of *J. curcas rbcL*, the *published* primers were able to group closely related species of *Jatropha* within the first 5 of 100 BLAST results. I expected that the BLAST analysis would show similar species within the *Jatropha* genus because of the conserved evolution of *rbcL.* Further efforts to amplify DNA using this primer pair with the Clontech kit were not needed as clear results were obtained fom the first PCR run using the Phire Plant kit (figure 6).

Analysis of modified rbcL primers against Jatropha curcas

The *modified* primers were a combination of the forward primer of *rbcL* from *published* primers and the reverse primer from pair 4 of our *designed* primers for *rbcL*. This was done to retrieve a full length gene sequence of *rbcL* for *J. curcas* using a single set of primers. This provided good results with very few gaps, coverage of the gene was above 95%, and nucleotide identity matches greater than 98% throughout the sequence. This region was also easily amplified without modification of the PCR conditions. Alignment of the forward and reverse primer worked well, with the forward primer giving 1,065 base pairs and the reverse giving 914 base pairs. The top 5 sequences matching my results were *J. curcas, J. capensis,* and *J. Intergerrima*, which suggests this gene is closely related to other species as *rbcL* evolves slowly. The amplification success rate and sequence quality using the *modified* primers are similar to what I found with the *published* primers showing the sequences obtained using the described primers were accurate. Further efforts to amplify DNA using primer pair 4 with the Clontech kit was not needed as successful amplification was seen during the first PCR run using the Phire Plant kit.

Amplifcation of rbcL and matK sequences of Lactuca sativa

Sequence amplification was obtained with the 2nd half of *Lactuca sativa rbcL* gene (primer pairs 3 and 4) by doubling the amount of template DNA added to PCR preparation. The second half of the *L. sativa* sequences was sent to the University of Washington sequencing facility by me and followed the same procedure as previously described. The sequence quality was good with the exception of the number of gaps. The number of gaps between the top 5 sequences showed in similar regions as seen with *J. curcas* suggesting that the gaps represent conserved regions. The sequence also produced a clear chromatogram with distinct single peaks for each nucleotide position when viewed with Finch TV.

Problems with amplification of the first half of the *Lactuca sativa* gene of *rbcL* and *matK* could be because the primers were contaminated during previous experiments. A new batch of primers was made to decrease the probability of contamination. Even though a new batch of primers was made, primer pairs 1 and 2 did not produce amplicons. This is contrary to the successful amplifcation of the *rbcL* gene sequence using a new working stock of the *designed* primers against *J.* curcas (primer pair 3). One might question the quality of template DNA used as the sample came from the local grocery store in Cheney, WA. Prolonged shelf storage and the associated storage conditions could have caused a decrease in the quality of DNA available. The *published matK* primers used against *L. sativa* in this study did not amplify the *matK* gene either. A PCR gradient for the reannealing temperature of *matK* was used but did not provide amplicons that could be used for sequencing. Unsuccessful amplification in the first half of the *rbcL* gene and the entire *matK* gene may be due to primer design. It is important to note that the control primers provided for by the Phire Plant kit gave correct fragment sizes during all amplification attempts indicating that the reagents from the kit were unspoiled.

Amplification of rbcL and matK sequences of Jatropha podagrica

Using the *designed* and *published* primers against *J. podagrica*, no amplification was obtained. A number of PCR trials were conducted, which yielded no amplicons even after new primers were made and a PCR gradient spanning a reannealing temperature range of 16 degrees celsius was used with *designed* primers for *matK.* This suggests interference by an unknown component or a high level of dissimilarity between the primer and collected tissue samples. The *rbcL* gene sequence was also not amplified. One probable cause of the unsuccessful amplification of *rbcL* and *matK* gene is the large amount of latex seeping out of the punched hole of the leaf. Introduction of a foreign contaminant such as latex into the PCR reaction can make it unreliable or eliminate amplification. The presence of poly-phenolics together with the high polysaccharide content of the latex makes the isolation of high-quality intact nucleic acids problematic in addition to complicating extractions and interfering in enzymatic reactions such as PCR (Michiels 2003).

Amplification of rbcL and matK sequences of Arabidopsis thaliana

Primers were designed from the *Arabidopsis* sequence (gi|7525012). Therefore, the sequences of *rbcL* and *matK* should have been easily amplified from *A. thaliana*. No amplification was found using the *matK* primers. After the first three PCR trials using the *designed* primers of *rbcL* showed little amplification, touchdown PCR was conducted to find the optimal annealing temperature of *rbcL* and to reduce the rate of non-specific priming. Touchdown PCR (TDPCR) works by finding the optimal annealing temperature by gradually decreasing the annealing temperature by 1 degree from near the melting point (upper limit) down to lower temperatures well beyond traditional annealing temperatures. This method reduces the rate of non-specific priming by only allowing very specific base pairing between the primer and template. At lower temperatures, primers bind less specifically. This test produced no amplification across a 16 degree TD PCR range. One problem encountered during these experiments was that *Arabidopsis* leaves were fresh and small, making it very difficult to prepare a leaf disc with the hole punch. The leaves tended to rip rather than produce a clean hole punch, which increased the level of extracellular debris and made amplification more difficult. To overcome this problem, I extended the time between harvesting and experimentation by drying the leaves enough for an accurate collection of template and to reduce starch content. This modification continually showed no amplification. Using the Plant Phire kit, the protocol for retrieving leaf tissue did not recommend punching through leaf veins. *Arabidopsis* leaves I used were not large and therefore avoiding veins during tissue retrieval were difficult. Since *Arabidopsis* has been studied in depth and serves as a model organisms in plant studies, it is unclear why amplification was not seen at higher rates and concentration. One way to avoid problems with obtaining appropiate sized leaf disc would be to wait for a longer period after planting seeds to let the leaves grow longer.

Designed, published, and modified primers using the Clontech kit

When using the Phire Plant kit, the success of amplification in *J. curcas rbcl* gene increased with the modification of newly made primers. The second half of the *L. sativa* gene amplification was easily attainable when the Plant Phire kit was used. The Plant Phire kit was used for experiments over a 2 year span and towards the end of the experiment the kit produced no amplifications. To obtain results using a different kit the *designed*, *published*, and *modified* primers were used with the new Clontech kit (table 9). *Designed*, *published*, and *modified* primers were run against *L. sativa* and *A. thaliana*. *Jatropha curcas* was not tested because sequence data was already collected and seasonal leaf deterioration made high quality DNA retrieval difficult. *Jatropha podagrica* contained a high amount of latex which was introduced into the PCR reaction and produced no amplification. Results showed smearing or no amplification in *L. sativa* and *A. thaliana*. Possible causes for unsuccessful amplification can be due to the need of different thermocycling conditions. The different Clontech protocol thermocycling conditions could have affected the annealing temperature of the primers. Attempts to find the optimal annealing temperature through touchdown PCR did not produce any changes in amplification. We suggest that the smearing is connected with the extracellular debris from the collection of template DNA of *Arabidopsis* and the low quality DNA of *L. sativa* obtained from the grocery store.

V. Conclusion

DNA barcoding provides a way for researchers to identify unknown species using a standard region of DNA. The Consortium for the Barcode of Life is proposing two genes to be used for DNA Barcode studies. The first of two genes, *rbcL*, is responsible for the fixation of carbon dioxide in the Calvin cycle. The second proposed gene region is *matK,* which is involved with group II splicing introns. This study showed that *rbcL* can be successfully amplified and sequenced with little to no modification in the case of *J. curcas.* Amplification of the *rbcL* gene was successful in 44% of attempts for one out of four pairs of the *designed* primers and the *published* primers worked the first time. Multiple runs of the *designed rbcL* primers eventually amplified all of our desired regions. The 3' end of the *rbcL* gene of *L. sativa* was amplifiable using the *designed* primers. The *designed* primers used against all other plants in this study did not produce any amplicons with sufficient amounts of DNA to move forward to sequencing. Though the *published* primers worked well with *J. curcas*, there was no amplification in *J. podagrica, L. sativa, and A. thaliana* using the same *published* primers, and in all cases when amplification was not achieved, positive control primers gave the correct product size. The *rbcL* gene is a good candidate for DNA barcoding and is capable of identifying unknown species or to place it within a family or a genus as shown by BLAST results, but until amplification across a variety of plants can be achieved without multiple runs, more work needs to be done in primer design.

One problem with the concept of a universal primer is in the process of tissue sampling. In this study, *J. podagrica* plant produced so much latex that clean uncontaminated PCR reactions were nearly impossible to achieve; the one successful amplification produced a low concentration of DNA. With the amount of classified and unclassified plants in the world exhibiting various morphologies, a more accurate way of harvesting leaves and preparation of leaf discs needs to be developed.

Maturase K was not amplified in any of the tests conducted. Hilu and Liang (1997) have reported that *matK* primers are readily amplifiable and can be useful to create a phylogenetic tree for the *matK* gene sequences. Although 100% amplification success is not realistic, amplification should be obtainable with ease and accuracy. In the current study, a PCR gradient test and touchdown PCR did nothing to alleviate amplification issues. In some studies, the *matK* gene showed that levels of amplification are lower than other proposed barcodes such as *rbcL* (CBOL Plant Working Group 2009). The *matK* gene produced inconclusive evidence as to its applicability as a DNA barcode; therefore, we suggest that a different gene be considered for the two-locus combination barcode.

Future Research

The current study shows that *rbcL* can be helpful as a barcode because of its good discriminatory value and sequence quality. Although the experiment only worked in *J. curcas*, physical characteristics of the leaves of *J. podagrica* and *A. thaliana* played a big part in the lack of success in amplifying the gene regions *rbcL* and *matK*. I suggest that high quality primers be designed and verified by more researchers to support the use of the *rbcL* gene region as a standard DNA Barcode. Though the *rbcL* gene shows the most promise for DNA barcoding in this study, another gene region other than *matK* should be considered. The continued search for a DNA barcode for land plants will result in larger databases of referenced sequences which will support the standardizing of a DNA barcode. In addition to modifications in the process of collection and extraction of high quality DNA, different gene regions should be evaluated as a dual locus barcode to be used with the *rbcL* gene.

VI. References

0. Arabidopsis Genome Initiative. (2000). Analysis of the Genome Sequence of the Flowering Plant Arabidopsis thaliana. Nature. 408: 796-815.

1. Barta, P. (2007). Jatropha Gains Steam in Global Race for Biofuels. The Wall Street Journal. Pg A1.

2. Becker K., Makkar P. (2008). Jatropha curcas: A potential source for tomorrow's oil and biodiesel. Lipid Technology. 20(5): 104-107.

3. CBOL Plant Working Group (2009). A DNA Barcode for Land Plants. PNAS. 106(31): 12794-12797.

4. Chase M.W., et al. (2007). A Proposal for a Standardised Protocol to Barcode All Land Plants. Taxon. 56(2): 295-299.

5. Chase M.W., Fay M.F. (2009). Barcoding of Plants and Fungi. Ecology. 325: 682-683.

6. Clegg M.T. (1993). Chloroplast Gene Sequences and the Study of Plant Evolution. PNAS. 90: 363-367.

7. Cooke E. (2009). Jatropha Biofuels; Miracle Plant or Economic Hazard? E555 Energy & the Environment. Indiana Universitiy-Bloomington.

8. Cullis C.A., et al. (2005). Mechanisms and control of rapid genomic changes in flax. Annals of Botany. 95: 201-206.

9. Cullis C.A., et al. (2009). Transfer of genetic material between the chloroplast and nucleus: how is it related to stress in plants? Annals of Botany. 103: 625-633.

10. Daniell, H., et al. (2004). Molecular Biology of Biotechnology of Plant Organelles. 3: 443-490. DOI: 10.1007/978-14020-3166-3_16.

11. Deghan B., Webster G.L. (1979). Morphology and infrageneric relationships of the genus Jatropha (Euphorbiaceae). Univ Calif Publ Bot. 74: 1-73.

12. Demkin, V. (2009). Bioinformatic Analysis of Bacterial Nucleotide Sequences Stored in GenBank. Molecular Genetics, Microbiology, and Virology. 24(2): 87-90.

13. Diekmann K., et al. (2008). An Optimized Chloroplast DNA Extraction Protocol for Grasses (Poaceae) Proves Suitable for Whole Plastid Genome Sequencing and SNP Detection. PLoS ONE. 3(7): 1-5.

14. Fankhauser, D.B. (2007). Isolation of Chloroplasts by Differential Centrifugation. University of Cincinnati Clermont College. Pg 1-4.

15. Fazekas A.J., et al. (2008). Multiple multilocus DNA barcodes from the plastid genome discriminate plant species equally well. PLoS One. 3: e2802. doi:10.1371/journal.pone. 0002802.

16. Greenacrescostarica.com. Retrieved March 2012.

17. Gubitz G.M., et al. (1999). Exploitation of the tropical oil seed plant Jatropha curcas L. Bioresource Technology. 67: 73–82.

18. Hasebe M., et al. (1994). rbcL gene sequences provide evidence for the evolutionary lineages of leptosporangiate ferns. PNAS. 91(12): 5730-5734.

19. Hebert P.D.N., et al. (2003) Barcoding animal life: cytochrome c oxidase subunit 1 divergences among closely related species. Proceedings of the Royal Society B: Biological Sciences. 270: S96–S99.

20. Hebert P. D. N., Stoeckle M. Y., Zemlak T. S. & Francis C. M. (2004). Identification of birds through DNA barcodes. PloS Biology. 2: 1657–1663.

21. Heller J. (1996). Physic nut Jatropha curcas L. Promoting the conservation and use of underutilized and neglected crops. Institute of Plant Genetics and Crop Plant Research. 6-42.

22. Hilu K., Liang H. (1997). The matk gene: sequence variation and application in plant systematics. Am. J. Bot. 84(6): 830.

23. Holmquist R. (1983). Transitions and transversion in evolutionary descent: an approach to understanding. Journal of Molecular Evolution. 19: 134-144.

24. Johnson L.A. and Soltis D.E. (1994). matK DNA sequences and phylogenetic reconstruction in Saxifragaceae s. str. Systematic Botany. 19(1): 143-156.

25. Kashdan MA., et al. (1980). Nucleotide Sequence of a Spinach Chloroplast Threonine tRNA. The Journal of Biological Chemistry. 255: 8831-8835.

26. King A.J., et al. (2009). Potential of Jatropha curcas as a source of renewable oil and animal feed. Journal of Experimental Botany. 60(10): 2897-2905.

27. Kress J.W., Erickson D.L. (2007). A Two-Locus Global DNA Barcode for Land Plants: The Coding rbcL Gene Complements the Non-Coding trnH-psbA Spacer Region. PLoS One. 2 (6): e508. doi:10.1371/journal.pone.0000508.

28. Kumar R.S. (2008). Genetic diversity among Jatropha species as revealed by RAPD markers. Genet Resour Crop Evol. 55: 803-809.

29. Lahaye R., et al. (2008). DNA Barcoding the Floras of Biodiversity Hotspots. PNAS. 105(8): 2923-2928.

30. Makkar P., Becker K. (2009). Jatropha curcas, a promising crop for the generation of biodiesel and value-added coproducts. European Journal of Lipid Science and Technology. 111(8): 773-787.

31. Martinez-Herrera J., et al. (2006). Chemical composition, toxic/antimetabolic constituents, and effects of different treatments on their levels, in four provenances of Jatropha curcas L. from Mexico. Food Chemistry. 96: 80–89.

32. Michiels A., et al. (2003). Extraction of high-quality genomic DNA from latex-containing plants. Analytical Biochemistry. 315: 85-89.

33. Moritz C., and C Cicero. (2004). DNA barcoding: Promise and pitfalls. PLoS Biol. 2: 1529–1531.

34. Neuhaus H., Link G. (1987). The chloroplast tRNALYS(UUU) gene from mustard (Sinapis alba) contains a class II intron potentially coding for a maturase-related polypeptide. Current Genetics. 11: 251–257.

35. Olmstead R.G. Palmer J.D. (1994). Chloroplast DNA systematics: a review of methods and data analysis. American Journal of Botany. 81(9): 1205-1224.

36. Palmer J.D., Herbon L.A. (1987). Unicircular Structure of the Brassica campestris Mitochondrial Genome. Current Genetics. 11: 565-570.

37. Palmer J.D., Herbon L.A. (1988). Plant Mitochondrial DNA Evolves Rapidly in Structure, but Slowly in Sequence. Journal of Molecular Evolution. 28: 87-97.

38. Palmer J.D. (1991) Plastid Chromosomes: Structure and Evolution. Cell Culture and Somatic Cell Genetics of Plants. 7A: 5-44.

39. Palmer J.D., et al. (2000). Dynamic Evolution of Plant Mitochondrial Genomes: Mobile Genes and Introns and Highly Variable Mutation Rates. PNAS. 97(13): 6960-6966.
40. Pamidimarri S., et al. (2009). Genetic divergence and phylogentic analysis of genus Jatropha based on nuclear ribosomal DNA ITS sequence. Mol Biol Rep. 36: 1929-1935.
41. Pamidimarri S., et al. (2009). Molecular characterization and indentification of markers for toxic and non-toxic varieties of Jatropha curcas L. using RAPD, AFLP and SSR markers. Mol. Biol Rep. 36: 1357-1364.
42. Piredda R., et al. (2011). Prospects of barcoding the Italian wild dendroflora: oaks reveal severe limitations to tracking species identity. Molecular Ecology Resources. 11(1): 72-83.
43. Fazekas AJ., et al. (2008). Multiple Multilocus DNA Barcodes from the Plastid Genome Discriminate Plant Species Equally Well. PLoS ONE. 3(7): 1-12.
44. Stit M., et al. (2010). Arabidopsis and Primary Photosynthetic Metabolism - More than the Icing on the Cake. The Plant Journal. 61: 1067-1091.
45. Sugita M., et al. (1985). Tobacco chloroplast tRNALys(UUU) gene contains a 2.5-kilobase-pair intron: an open reading frame and a conserved boundary sequence in the intron. PNAS. 82: 3557– 3561.
46. Tabita R.F., et al. (2007). Functoin, Structure, and Evolution of the RubiCO-Like Proteins and Their RubisCO Homologs. Microbiology and Molecular Biology Reviews. 71(4): 576-599.
47. Valentini A., et al. (2008). DNA barcoding for ecologists. Trends Ecol Evol. 24: 110–117.
48. Veen Van J., Schucker C., Kennedy A. (2010). RubisCO gene comparison between three Lactuca sativa cultivars. Eastern Washington University. (Poster).

49. Wicke S. (2009). Universal primers for the amplification of the trnK/matK region in land plants. Anales Jard Bot Madrid. 66(2): 285-288.

50. Youngs H. (2012). Development of feedstocks for cellulosic biofuels. F1000 Biol Rep. 4: 1-11.

51. Zhang L., et al. (2009). Chemical Constituents of the Plants from Genus Jatropha. Chemistry and Biodiversity. 6(12): 2166-2183.

52. Zurawski G. (1984). Evolution of Higher-Plant Chloroplast DNA-Encoded Genes: implications for Structure-Function and Phylogenetic Studies. Annual Review of Plant Physiology. 38: 391-418.

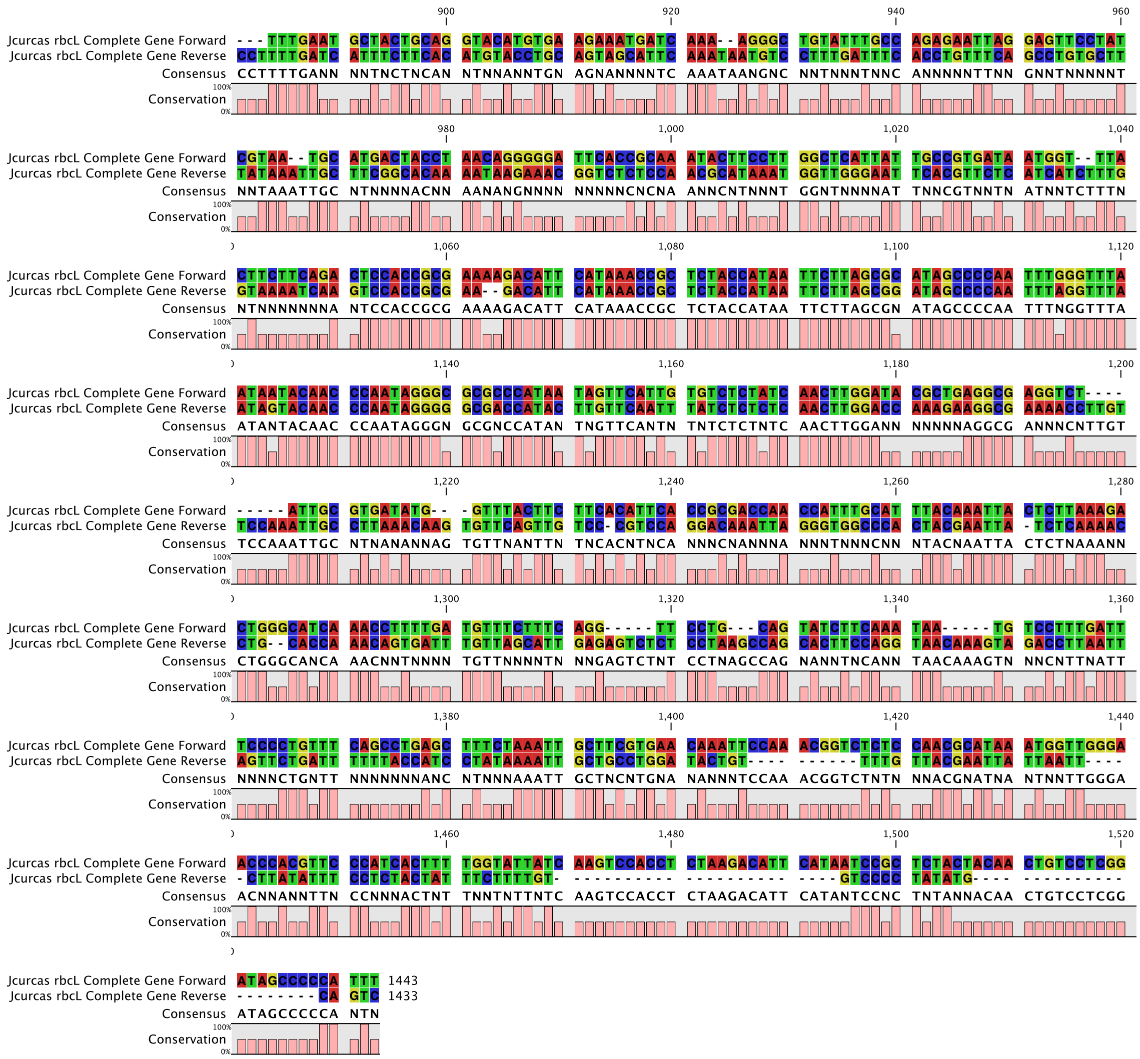
Appendix 1

(*Jatropha curcas rbcL* gene using *designed* primers)

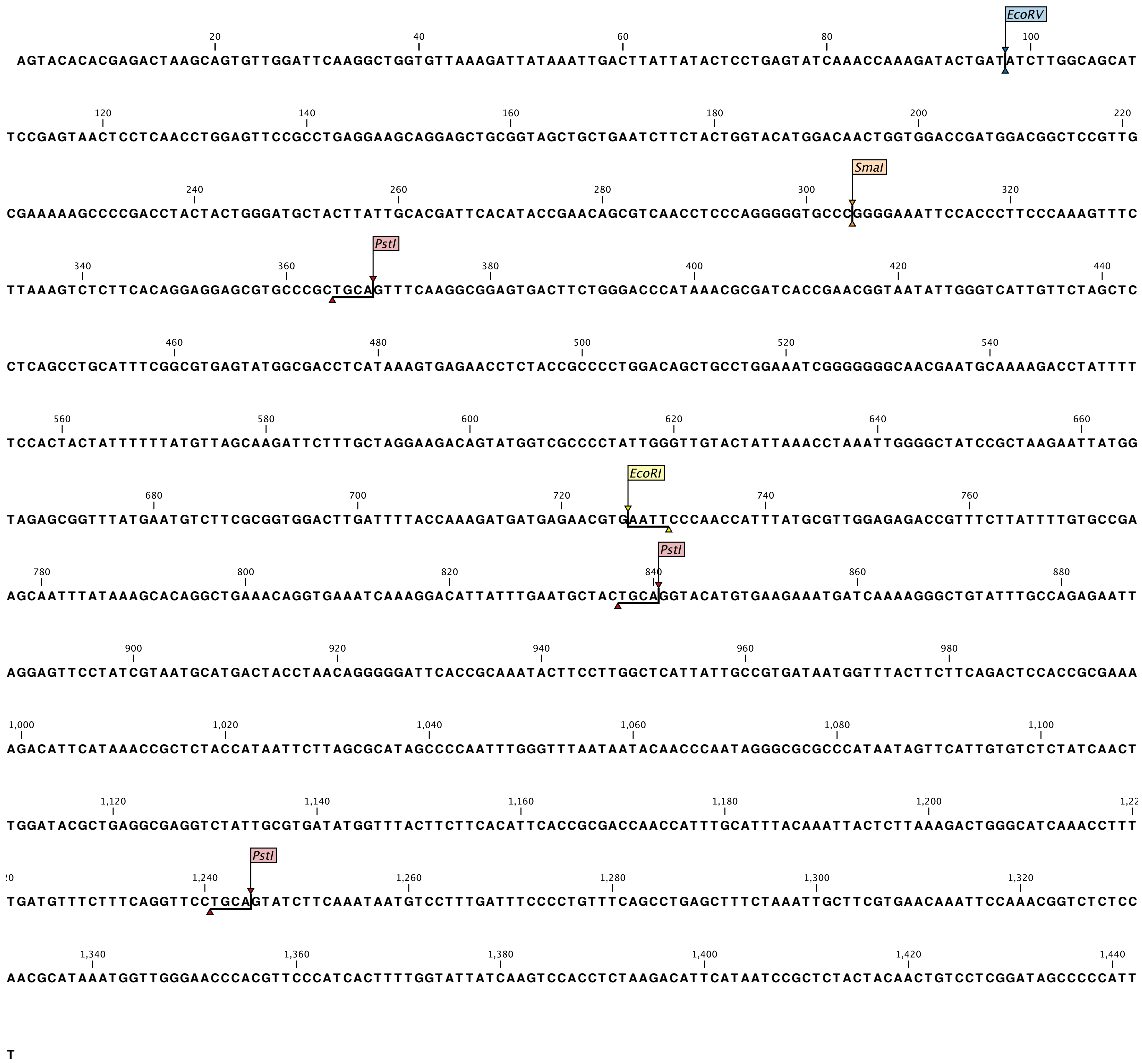
I. Aligned Sequences

II. Restriction Sites

J. curcas rbcL gene alignment using designed primers


900 920 940 960
Jcurcas rbcL Complete Gene Forward ---TTTGAAT GCTACTGCAG GTACATGTGA AGAAATGATC AAA--AGGGC TGTATTTGCC AGAGAATTAG GAGTTCCTAT
Jcurcas rbcL Complete Gene Reverse CCTTTTGATC ATTTCTTCAC ATGTACCTGC AGTAGCATTC AAATAATGTC CTTTGATTTC ACCTGTTTCA GCCTGTGCTT
Consensus CCTTTTGANN NNTNCTNCAN NTNNANNTGN AGNANNNNTC AAATAANGNC NNTNNNTNNC ANNNNNTTNN GNNTNNNNNT
Conservation 100% 0%
980 1,000 1,020 1,040
Jcurcas rbcL Complete Gene Forward CGTAA--TGC ATGACTACCT AACAGGGGGA TTCACCGCAA ATACTTCCTT GGCTCATTAT TGCCGTGATA ATGGT--TTA
Jcurcas rbcL Complete Gene Reverse TATAAATTGC TTCGGCACAA AATAAGAAAC GGTCTCTCCA ACGCATAAAT GGTTGGGAAT TCACGTTCTC ATCATCTTTG
Consensus NNTAAATTGC NTNNNNACNN AANANGNNNN NNNNNCNCNA ANNCNTNNNT GGNTNNNNAT TNNCGTNNTN ATNNTCTTTN
Conservation 100% 0%
1,060 1,080 1,100 1,120
Jcurcas rbcL Complete Gene Forward CTTCTTCAGA CTCCACCGCG AAAAGACATT CATAAACCGC TCTACCATAA TTCTTAGCGC ATAGCCCCAA TTTGGGTTTA
Jcurcas rbcL Complete Gene Reverse GTAAAATCAA GTCCACCGCG AA--GACATT CATAAACCGC TCTACCATAA TTCTTAGCGG ATAGCCCCAA TTTAGGTTTA
Consensus NTNNNNNNNA NTCCACCGCG AAAAGACATT CATAAACCGC TCTACCATAA TTCTTAGCGN ATAGCCCCAA TTTNGGTTTA
Conservation 100% 0%
1,140 1,160 1,180 1,200
Jcurcas rbcL Complete Gene Forward ATAATACAAC CCAATAGGGC GCGCCCATAA TAGTTCATTG TGTCTCTATC AACTTGGATA CGCTGAGGCG AGGTCT----
Jcurcas rbcL Complete Gene Reverse ATAGTACAAC CCAATAGGGG GCGACCATAC TTGTTCAATT TATCTCTCTC AACTTGGACC AAAGAAGGCG AAAACCTTGT
Consensus ATANTACAAC CCAATAGGGN GCGNCCATAN TNGTTCANTN TNTCTCTNTC AACTTGGANN NNNNNAGGCG ANNNCNTTGT
Conservation 100% 0%
1,220 1,240 1,260 1,280
Jcurcas rbcL Complete Gene Forward -----ATTGC GTGATATG-- -GTTTACTTC TTCACATTCA CCGCGACCAA CCATTTGCAT TTACAAATTA CTCTTAAAGA
Jcurcas rbcL Complete Gene Reverse TCCAAATTGC CTTAAACAAG TGTTCAGTTG TCC-CGTCCA GGACAAATTA GGGTGGCCCA CTACGAATTA -TCTCAAAAC
Consensus TCCAAATTGC NTNANANNAG TGTTNANTTN TNCACNTNCA NNNCNANNNA NNNTNNNCNN NTACNAATTA CTCTNAAANN
Conservation 100% 0%
1,300 1,320 1,340 1,360
Jcurcas rbcL Complete Gene Forward CTGGGCATCA AACCTTTTGA TGTTTCTTTC AGG-----TT CCTG---CAG TATCTTCAAA TAA-----TG TCCTTTGATT
Jcurcas rbcL Complete Gene Reverse CTG--CACCA AACAGTGATT TGTTAGCATT GAGAGTCTCT CCTAAGCCAG CACTTCCAGG TAACAAAGTA GACCTTAATT
Consensus CTGGGCANCA AACNNTNNNN TGTTNNNNTN NNGAGTCTNT CCTNAGCCAG NANNTNCANN TAACAAAGTN NNCNTTNATT
Conservation 100% 0%
1,380 1,400 1,420 1,440
Jcurcas rbcL Complete Gene Forward TCCCCTGTTT CAGCCTGAGC TTTCTAAATT GCTTCGTGAA CAAATTCCAA ACGGTCTCTC CAACGCATAA ATGGTTGGGA
Jcurcas rbcL Complete Gene Reverse AGTTCTGATT TTTTACCATC CTATAAAATT GCTGCCTGGA TACTGT---- ------TTTG TTACGAATTA TTAATT----
Consensus NNNNCTGNTT NNNNNNNANC NTNNNAAATT GCTNCNTGNA NANNNTCCAA ACGGTCTNTN NNACGNATNA NTNNTTGGGA
Conservation 100% 0%
1,460 1,480 1,500 1,520
Jcurcas rbcL Complete Gene Forward ACCCACGTTC CCATCACTTT TGGTATTATC AAGTCCACCT CTAAGACATT CATAATCCGC TCTACTACAA CTGTCCTCGG
Jcurcas rbcL Complete Gene Reverse -CTTATATTT CCTCTACTAT TTCTTTTTGT ---------- ---------- ----GTCCCC TATATG---- ----------
Consensus ACNNANNTTN CCNNNACTNT TNNTNTTNTC AAGTCCACCT CTAAGACATT CATANTCCNC TNTANNACAA CTGTCCTCGG
Conservation 100% 0%
Jcurcas rbcL Complete Gene Forward ATAGCCCCCA TTT 1443
Jcurcas rbcL Complete Gene Reverse --------CA GTC 1433
Consensus ATAGCCCCCA NTN
Conservation 100% 0%

J. curcas rbcL gene restriction sites using designed primers (forward strand)



1

J. curcas rbcL gene restricton sites using designed primers (reverse strand)

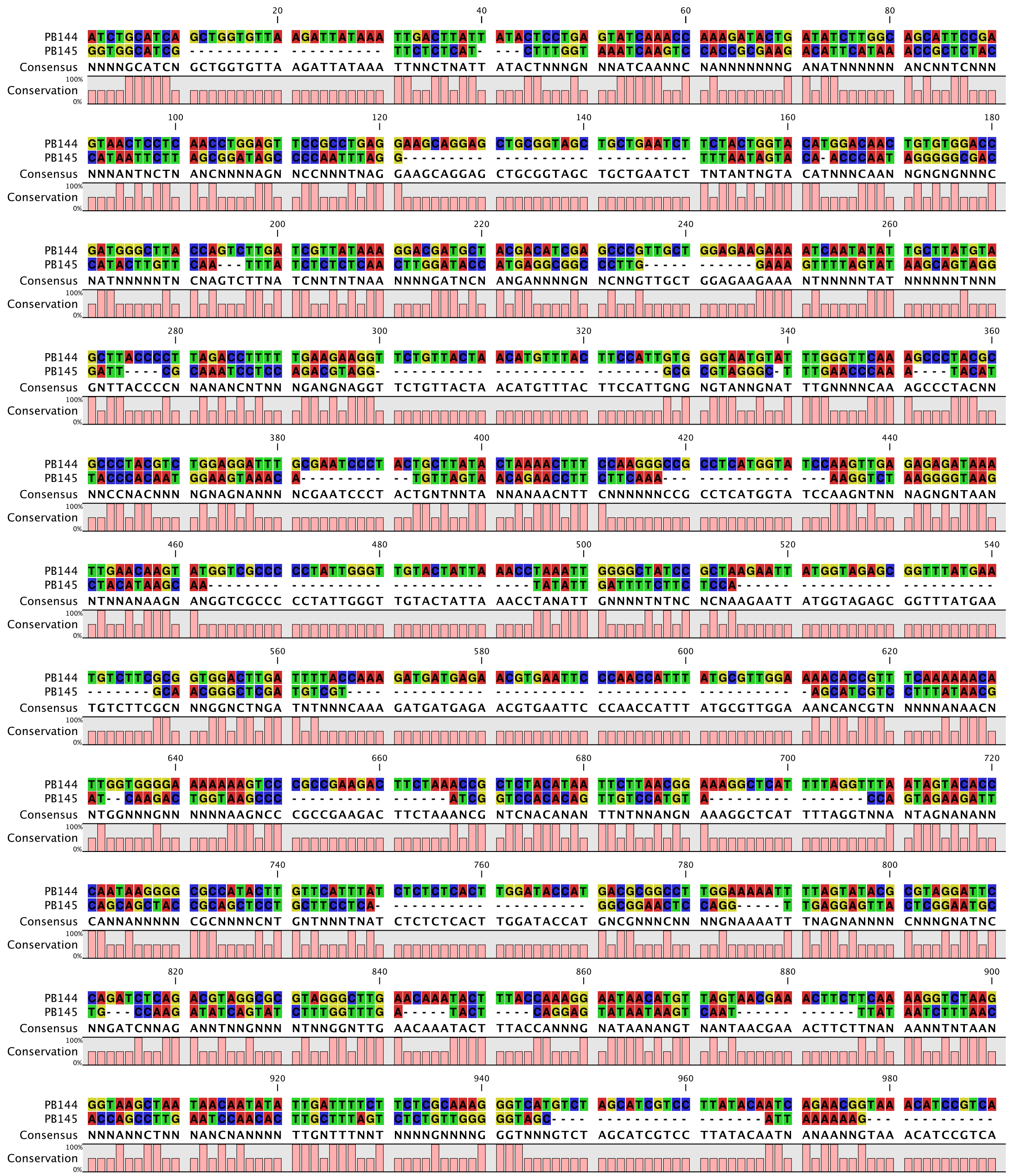
```
                                   20         40         60         80
Jcurcas rbcL Complete Gene Reverse CAAAAATTTGCCCAGTTACTCTTCCTCAGTGCTGTATGCCCGCTGCCGTAACAAAGTTAAAATGACTTATTATACCCCTGAAAAGCAG
                                   100        120        140        160
                                                                      SalI
Jcurcas rbcL Complete Gene Reverse ACCAACGATACTGAGATATTGGCAGCATTCCGAGTATCTCCTCAGCCTGTAGTTCCGCCTGAGGAAGGTCGACCTGCATAAACTGCTG
                                   180        200        220        240        260
                                                          BamHI
Jcurcas rbcL Complete Gene Reverse AATCTTCTACCACCCCATGAACAACTGGCTGGTCCGATGGGCGGGATCCGATCTGGCAAAGACCGTCATCTCCAGCACTATTATTATT
                                   280        300        320        340
Jcurcas rbcL Complete Gene Reverse GTGAACCCGGTTTATGGCTACATCATGTTGGACCAACCTATGGGTGTCCTGGAGGGTTTTCCTTCTCCCCTGAGTGGGAAGCCTATCG
                                   360        380        400        420        440
Jcurcas rbcL Complete Gene Reverse GGACCCACCCGCTAAACTTCTTAGCGGAAGCCCATTTAGGTTTAATAGTACAACCCAATAGGGGGCGACCATACTTGTTCAATTTATC
                                   0          460        480        500        520
Jcurcas rbcL Complete Gene Reverse TCTCTCAACTTGGATACCATGAGGCGGCCCTTGGAAAGTTTTAGTATAAGCAGTAGGGATTCGCAAATCCTCCAGACGTAGGGCGCGT
                                   540        560        580        600
Jcurcas rbcL Complete Gene Reverse AGGGCTTTGAACCCAAATACATTACCCACAATGGAAGTAAACATGTTAGTAACCGAACCTTCTTCAAAAAGGTCTAAGGGGTAAACTA
                                   620        640        660        680        700
Jcurcas rbcL Complete Gene Reverse TCTAAGTTGTGAACCGGGTCTATGGCTACTTCGTGCTGGGCCACCCTATGTGTGTCCTGGAGGGCTACACCGTCTCCCTGTGTGGTAA
                                   720        740        760        780
Jcurcas rbcL Complete Gene Reverse GCCTTTCGGTGCCCAGGCTCGACGGAAACCACTCTTTCACCCTGCCATTCCCTACGCATATGAGCCAGGAGTATTTGCGGTGAATCCC
                                   800        820        840        860        88
                                                                                         P
Jcurcas rbcL Complete Gene Reverse CCTGTTAGGTAGTCATGCATTACGATAGGAACTCCTAATTCTCTGGCAAATACAGCCCTTTTGATCATTTCTTCACATGTACCTGCAG
                                   0          900        920        940        960
                                   stI
Jcurcas rbcL Complete Gene Reverse TAGCATTCAAATAATGTCCTTTGATTTCACCTGTTTCAGCCTGTGCTTTATAAATTGCTTCGGCACAAAATAAGAAACGGTCTCTCCA
                                   980        1,000      1,020      1,040
                                                 EcoRI
Jcurcas rbcL Complete Gene Reverse ACGCATAAATGGTTGGGAATTCACGTTCTCATCATCTTTGGTAAAATCAAGTCCACCGCGAAGACATTCATAAACCGCTCTACCATAA
                                   1,060      1,080      1,100      1,120      1,140
Jcurcas rbcL Complete Gene Reverse TTCTTAGCGGATAGCCCCAATTTAGGTTTAATAGTACAACCCAATAGGGGGCGACCATACTTGTTCAATTTATCTCTCTCAACTTGGA
                                   1,160      1,180      1,200      1,220
Jcurcas rbcL Complete Gene Reverse CCAAAGAAGGCGAAAACCTTGTTCCAAATTGCCTTAAACAAGTGTTCAGTTGTCCCGTCCAGGACAAATTAGGGTGGCCCACTACGAA
```

Appendix 2

(*Jatropha curcas rbcL* gene using *published* primers)

III. Aligned Sequences

IV. Restriction Sites


PB144 ATCTGCATCA GCTGGTGTTA AGATTATAAA TTGACTTATT ATACTCCTGA GTATCAAACC AAAGATACTG ATATCTTGGC AGCATTCCGA
PB145 GGTGGCATCG ---------- ---------- TTGTCTCAT- ---CTTTGGT AAATCAAGTC CACCGCGAAG ACATTCATAA ACCGCTCTAC
Consensus NNNNGCATCN GCTGGTGTTA AGATTATAAA TTNNCTNATT ATACTNNNGN NNATCAANNC NANNNNNNNG ANATNNNNNN ANCNNTCNNN
Conservation
PB144 GTAACTCCTC AACCTGGAGT TCCGCCTGAG GAAGCAGGAG CTGCGGTAGC TGCTGAATCT TGTACTGGTA CATGGACAAC TGTGTGGACC
PB145 CATAATTCTT AGCGGATAGC CCCAATTTAG G--------- ---------- ---------- TTTAATAGTA CA-ACCCAAT AGGGGGCGAC
Consensus NNNANTNCTN ANCNNNNAGN NCCNNNTNAG GAAGCAGGAG CTGCGGTAGC TGCTGAATCT TNTANTNGTA CATNNNCAAN NGNGNGNNNC
Conservation
PB144 GATGGGCTTA CCAGTCTTGA TCGTTATAAA GGACGATGCT ACGACATCGA GCCCGTTGCT GGAGAAGAAA ATCAATATAT TGCTTATGTA
PB145 CATACTTGTT CAA---TTTA TCTCTCTCAA CTTGGATACC ATGAGGCGGC CCTTG----- ------GAAA GTTTTAGTAT AAGCAGTAGG
Consensus NATNNNNNTN CNAGTCTTNA TCNNTNTNAA NNNNGATNCN ANGANNNNGN NCNNGTTGCT GGAGAAGAAA NTNNNNNTAT NNNNNNTNNN
Conservation
PB144 GCTTACCCCT TAGACCTTTT TGAAGAAGGT TCTGTTACTA ACATGTTTAC TTCCATTGTG GGTAATGTAT TTGGGTTCAA AGCCCTACGC
PB145 GATT----CG CAAATCCTCC AGACGTAGG- ---------- ---------- -------GCG CGTAGGGC-T TTGAACCCAA A----TACAT
Consensus GNTTACCCCN NANANCNTNN NGANGNAGGT TCTGTTACTA ACATGTTTAC TTCCATTGNG NGTANNGNAT TTGNNNNCAA AGCCCTACNN
Conservation
PB144 GCCCTACGTC TGGAGGATTT GCGAATCCCT ACTGCTTATA CTAAAACTTT CCAAGGGCCG CCTCATGGTA TCCAAGTTGA GAGAGATAAA
PB145 TACCCACAAT GGAAGTAAAC A--------- --TGTTAGTA ACAGAACCTT CTTCAAA--- ---------- ---AAGGTCT AAGGGGTAAG
Consensus NNCCNACNNN NGNAGNANNN NCGAATCCCT ACTGNTNNTA NNANAACNTT CNNNNNNCCG CCTCATGGTA TCCAAGNTNN NAGNGNTAAN
Conservation
PB144 TTGAACAAGT ATGGTCGCCC CCTATTGGGT TGTACTATTA AACCTAAATT GGGGCTATCC GCTAAGAATT ATGGTAGAGC GGTTTATGAA
PB145 CTACATAAGC AA-------- ---------- ---------- ----TATATT GATTTTCTTC TCCA------ ---------- ----------
Consensus NTNNANAAGN ANGGTCGCCC CCTATTGGGT TGTACTATTA AACCTANATT GNNNNTNTNC NCNAAGAATT ATGGTAGAGC GGTTTATGAA
Conservation
PB144 TGTCTTCGCG GTGGACTTGA TTTTACCAAA GATGATGAGA ACGTGAATTC CCAACCATTT ATGCGTTGGA AAACACCGTT TCAAAAAACA
PB145 -------GCA ACGGGCTCGA TGTCGT---- ---------- ---------- ---------- ---------- -AGCATCGTC CTTTATAACG
Consensus TGTCTTCGCN NNGGNCTNGA TNTNNNCAAA GATGATGAGA ACGTGAATTC CCAACCATTT ATGCGTTGGA AANCANCGTN NNNNANAACN
Conservation
PB144 TTGGTGGGGA AAAAAAGTCC CGCCGAAGAC TTCTAAACCG CTCTACATAA TTCTTAACGG AAAGGCTCAT TTTAGGTTTA ATAGTACACC
PB145 AT--CAAGAC TGGTAAGCCC ---------- ------ATCG GTCCACACAG TTGTCCATGT A--------- -------CCA GTAGAAGATT
Consensus NTGGNNNGNN NNNNAAGNCC CGCCGAAGAC TTCTAAANCG NTCNACANAN TTNTNNANGN AAAGGCTCAT TTTAGGTNNA NTAGNANANN
Conservation
PB144 CAATAAGGGG GGCCATACTT GTTCATTTAT CTCTCTCACT TGGATACCAT GACGCGGCCT TGGAAAAATT TTAGTATACG GGTAGGATTC
PB145 CAGCAGCTAC CGCAGCTCGT GCTTCCTCA- ---------- ---------- GGCGGAACTC CAGG-----T TGAGGAGTTA CTCGGAATGC
Consensus CANNANNNNN CGCNNNNCNT GNTNNNTNAT CTCTCTCACT TGGATACCAT GNCGNNNCNN NNGNAAAATT TNAGNANNNN CNNNGNATNC
Conservation
PB144 CAGATCTCAG ACGTAGGCGC GTAGGGCTTG AACAAATACT TTACCAAAGG AATAACATGT TAGTAACGAA ACTTCTTCAA AAGGTCTAAG
PB145 TG---CCAAG ATATCAGTAT GTTTGGTTTG A-----TACT ----CAGGAG TATAATAAGT CAAT------ ------TTAT AATCTTTAAC
Consensus NNGATCNNAG ANNTNNGNNN NTNNGGNTTG AACAAATACT TTACCANNNG NATAANANGT NANTAACGAA ACTTCTTNAN AANNTNTAAN
Conservation
PB144 GGTAAGCTAA TAACAATATA TTGATTTTCT TCTCGCAAAG GGTCATGTCT AGCATCGTCC TTATACAATC AGAACGGTAA ACATCCGTCA
PB145 ACCAGCCTTG AATCCAACAC TTGCTTTAGT CTCTGTTGGG GGTAGC---- ---------- -------ATT AAAAAAG--- ----------
Consensus NNNANNCTNN NANCNANNNN TTGNTTTNNT NNNNGNNNNG GGTNNNGTCT AGCATCGTCC TTATACAATN ANAANNGTAA ACATCCGTCA
Conservation

1

J. curcas rbcL gene restriction sites using published primers (forward strand)

```
                20        40        60        80        100
                                      EcoRV
PB144 ATCTGCATCAGCTGGTGTTAAGATTATAAATTGACTTATTATACTCCTGAGTATCAAACCAAAGATACTGATATCTTGGCAGCATTCCGAGTAACTCCTCAACCT
        120       140       160       180       200
PB144 GGAGTTCCGCCTGAGGAAGCAGGAGCTGCGGTAGCTGCTGAATCTTCTACTGGTACATGGACAACTGTGTGGACCGATGGGCTTACCAGTCTTGATCGTTATAAAG
      220       240       260       280       300
PB144 GACGATGCTACGACATCGAGCCCGTTGCTGGAGAAGAAAATCAATATATTGCTTATGTAGCTTACCCCTTAGACCTTTTTGAAGAAGGTTCTGTTACTAACATGTT
     320       340       360       380       400       420
PB144 TACTTCCATTGTGGGTAATGTATTTGGGTTCAAAGCCCTACGCGCCCTACGTCTGGAGGATTTGCGAATCCCTACTGCTTATACTAAAACTTTCCAAGGGCCGCCT
         440       460       480       500       520
PB144 CATGGTATCCAAGTTGAGAGAGATAAATTGAACAAGTATGGTCGCCCCCTATTGGGTTGTACTATTAAACCTAAATTGGGGCTATCCGCTAAGAATTATGGTAGAG
       540       560       580       600       620
                                           EcoRI
PB144 CGGTTTATGAATGTCTTCGCGGTGGACTTGATTTTACCAAAGATGATGAGAACGTGAATTCCCAACCATTTATGCGTTGGAAAACACCGTTTCAAAAAACATTGGT
      640       660       680       700       720       740
PB144 GGGGAAAAAAAGTCCCGCCGAAGACTTCTAAACCGCTCTACATAATTCTTAACGGAAAGGCTCATTTTAGGTTTAATAGTACACCCAATAAGGGGCGCCATACTTG
          760       780       800       820       840
                                                   BglII
PB144 TTCATTTATCTCTCTCACTTGGATACCATGACGCGGCCTTGGAAAAATTTTAGTATACGCGTAGGATTCCAGATCTCAGACGTAGGCGCGTAGGGCTTGAACAAAT
        860       880       900       920       940
PB144 ACTTTACCAAAGGAATAACATGTTAGTAACGAAACTTCTTCAAAAGGTCTAAGGGTAAGCTAATAACAATATATTGATTTTCTTCTCGCAAAGGGTCATGTCTAGC
      960       980       1,000
PB144 ATCGTCCTTATACAATCAGAACGGTAAACATCCGTCACATTTGTTCATGT
```

1

J. curcas rbcL gene restriction sites using published primers (reverse strand)

20 40 60 80 100

PB145 GGTGGCATCGTTCTCTCATCTTTGGTAAATCAAGTCCACCGCGAAGACATTCATAAACCGCTCTACCATAATTCTTAGCGGATAGCCCCAATTTAGGTTTAATAG

120 140 160 180 200

PB145 TACAACCCAATAGGGGGCGACCATACTTGTTCAATTTATCTCTCTCAACTTGGATACCATGAGGCGGCCCTTGGAAAGTTTTAGTATAAGCAGTAGGGATTCGCAA

220 240 260 280 300

PB145 ATCCTCCAGACGTAGGGCGCGTAGGGCTTTGAACCCAAATACATTACCCACAATGGAAGTAAACATGTTAGTAACAGAACCTTCTTCAAAAAGGTCTAAGGGGTAA

320 340 360 380 400 420

PB145 GCTACATAAGCAATATATTGATTTTCTTCTCCAGCAACGGGCTCGATGTCGTAGCATCGTCCTTTATAACGATCAAGACTGGTAAGCCCATCGGTCCACACAGTTG

440 460 480 500 520

EcoRV

PB145 TCCATGTACCAGTAGAAGATTCAGCAGCTACCGCAGCTCCTGCTTCCTCAGGCGGAACTCCAGGTTGAGGAGTTACTCGGAATGCTGCCAAGATATCAGTATCTTT

540 560 580 600 620

PB145 GGTTTGATACTCAGGAGTATAATAAGTCAATTTATAATCTTTAACACCAGCCTTGAATCCAACACTTGCTTTAGTCTCTGTTGGGGGTAGCATTAAAAAAG

1

Appendix 3

(*Jatropha curcas rbcL* gene using *modified* primers)

V. Aligned Sequences

VI. Restriction Sites

J. curcas rbcL gene alignment using modified primers

1

J. curcas rbcL gene restrictoin sites using modified primers (forward strand)

20 40 60 80 100

EcoRV

PB146 ATATGGATCACGCTGGTGTTAAGATTATAAATTGACTTATTATACTCCTGAGTATCAAACCAAAGATACTGATATCTTGGCAGCATTCCGAGTAACTCCTCAACC

120 140 160 180 200

PB146 TGGAGTTCCGCCTGAGGAAGCAGGAGCTGCGGTAGCTGCTGAATCTTCTACTGGTACATGGACAACTGTGTGGACCGATGGGCTTACCAGTCTTGATCGTTATAAA

220 240 260 280 300

PB146 GGACGATGCTACGACATCGAGCCCGTTGCTGGAGAAGAAAATCAATATATTGCTTATGTAGCTTACCCCTTAGACCTTTTTGAAGAAGGTTCTGTTACTAACATGT

320 340 360 380 400 420

PB146 TTACTTCCATTGTGGGTAATGTATTTGGGTTCAAAGCCCTACGCGCCCTACGTCTGGAGGATTTGCGAATCCCTACTGCTTATACTAAAACTTTCCAAGGGCCGCC

440 460 480 500 520

PB146 TCATGGTATCCAAGTTGAGAGAGATAAATTGAACAAGTATGGTCGCCCCCTATTGGGTTGTACTATTAAACCTAAATTGGGGCTATCCGCTAAGAATTATGGTAGA

540 560 580 600 620

EcoRI

PB146 GCGGTTTATGAATGTCTTCGCGGTGGACTTGATTTTACCAAAGATGATGAGAACGTGAATTCCCAACCATTTATGCGTTGGAGAGACCGTTTCTTATTTTGTGCCG

640 660 680 700 720 740

PstI

PB146 AAGCAATTTATAAAGCACAGGCTGAAACAGGTGAAATCAAAGGACATTATTTGAATGCTACTGCAGGTACATGTGAAGAAATGATCAAAAGGGCTGTATTTGCCAG

760 780 800 820 840

PB146 AGAATTAGGAGTTCCTATCGTAATGCATGACTACCTAACAGGGGGATTCACCGCAAATACTTCCTTGGCTCATTATTGCCGTGATAATGGTTTACTTCTTCACATT

860 880 900 920 940

PB146 CACCGCGCAATGCATGCAGTTATTGATAGACAGAAGAATCATGGTATGCATTTTCGTGTACTAGCTAAGGCGTTACGTTTGTCTGGTGGAGATCATATTCACGCTG

960 980 1,000 1,020 1,040 1,

PB146 GTACCGTAGTATGTAAACTTGAAAGGGAAAGAGACATCACTTTGGGCTTTGTTGATTTACTGCGTGATGATTTTATTGAAAAAGATCGAAGCCGCGGTATTTATTT

060

PB146 CACTCA

1

J. curcas rbcL gene restriction sites using modified primers (reverse strand)

```
                                        20                  40                  60                  80                  100
                                                              EcoRI
PB147  AAAAACGACGTCATTAGTGGTTACAAAGTATCCATCGCTTGGAATTCAAATTTAATCTCCTTCCATACTTCACAAGCAGCAGCTAGTTCAGGACTCCATTTGCTA
                  120                 140                 160                 180                 200
                                                                                 XbaI
PB147 GCCTCACGGATAATTTCATTACCCTCACGAGCAAGATCACGTCCTTCATTACGAGCTTGTACACATGCTTCTAGAGCTACTCGATTAGCTACGGCACCAGGTGCAT
            220                 240                 260                 280                 300
                                                                    BglII
PB147 TTCCCCAAGGGTGCCCTAAAGTTCCTCCACCGAATTGTAGTACGGAATCATCTCCAAAGATCTCGGTCAGAGCAGGCATATGCCAAACGTGAATACCTCCTGAAGC
      320                 340                 360                 380                 400                 420
PB147 TACAGGCAAAACACCCGGTAGAGAGACCCAATCTTGAGTGAAATAAATACCGCGGCTTCGATCTTTTTCAATAAAATCATCACGCAGTAAATCAACAAAGCCCAAA
                    440                 460                 480                 500                 520
PB147 GTGATGTCTCTTTCCCCTTCAAGTTTACCTACTACGGTACCAGCGTGAATATGATCTCCACCAGACAAACGTAACGCCTTAGCTAGTACACGAAAATGCATACCAT
              540                 560                 580                 600                 620
PB147 GATTCTTCTGTCTATCAATAACTGCATGCATTGCGCGGTGAATGTGAAGAAGTAAACCATTATCACGGCAATAATGAGCCAAGGAAGTATTTGCGGTGAATCCCCC
        640                 660                 680                 700                 720                 740
                                                                                               PstI
PB147 TGTTAGGTAGTCATGCATTACGATAGGAACTCCTAATTCTCTGGCAAATACAGCCCTTTTGATCATTTCTTCACATGTACCTGCAGTAGCATTCAAATAATGTCCT
                     760                 780                 800                 820                 840
                                                                                        EcoRI
PB147 TTGATTTCACCTGTTTCAGCCTGTGCTTTATAAATTGCTTCGGCACAAATAGAACGGTCTCTCCACGCATAAATGGTTGGGAATTCACGTTCTCATCATCTTTGGT
              860                 880                 900
PB147 AAATCAAGTCCACCGCGAAGACATCATAAACCGCTCTACCATAATTCTTAGCGGATAGCCCAATTAA
```

1

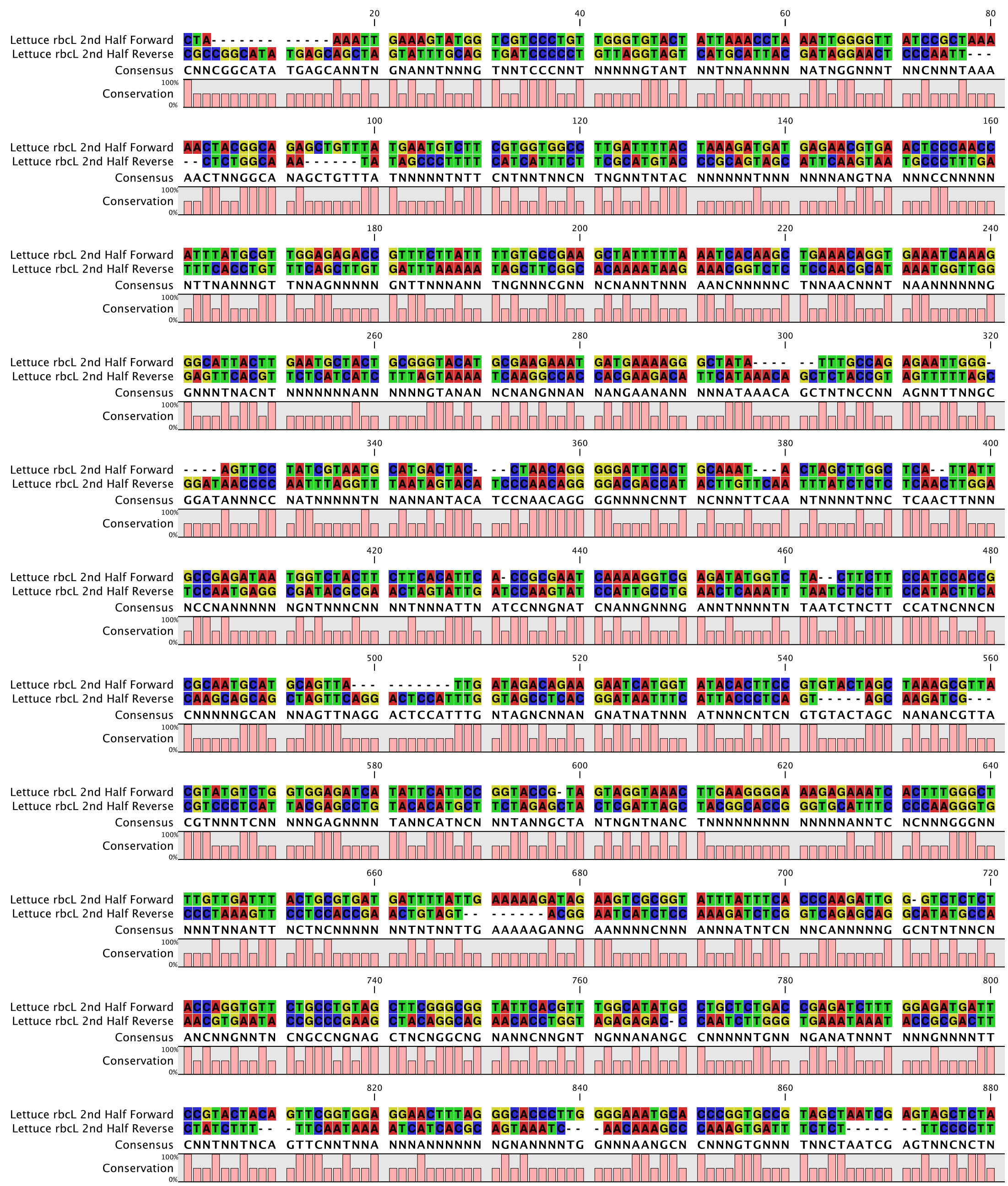
Appendix 4

(2nd half of *Latuca sativa rbcL* gene using *designed* primers)

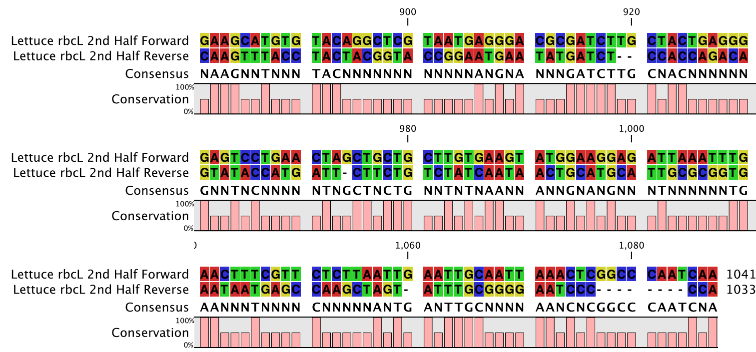
VII. Aligned Sequences

VIII. Restriction Sites

L. sativa rbcL gene alignment using designed primers



1



900
920
940
960
Lettuce rbcL 2nd Half Forward GAAGCATGTG TACAGGCTCG TAATGAGGGA CGCGATCTTG CTACTGAGGG TAATGAAATT ATCCGTGAGG CTACCAAATG
Lettuce rbcL 2nd Half Reverse CAAGTTTACC TACTACGGTA CCGGAATGAA TATGATCT-- CCACCAGACA TACGTAACGC TTTAGCTAGT ACACGGAAGT
Consensus NAAGNNTNNN TACNNNNNNN NNNNNANGNA NNNGATCTTG CNACNNNNNN TANNNAANNN NTNNGNNAGN NNACNNAANN
Conservation
100%
0%
980
1,000
1,020
1,040
Lettuce rbcL 2nd Half Forward GAGTCCTGAA CTAGCTGCTG CTTGTGAAGT ATGGAAGGAG ATTAAATTTG AGTTTCAGGC AATGGATACT TTGGATCAAT
Lettuce rbcL 2nd Half Reverse GTATACCATG ATT-CTTCTG TCTATCAATA ACTGCATGCA TTGCGCGGTG GATGTGAAGA AGTAGAC-CA TTATCTCGGC
Consensus GNNTNCNNNN NTNGCTNCTG NNTNTNAANN ANNGNANGNN NTNNNNNNTG NNTNTNANGN ANTNGANACN TTNNNTCNNN
Conservation
100%
0%
1,060
1,080
Lettuce rbcL 2nd Half Forward AACTTTCGTT CTCTTAATTG AATTGCAATT AAACTCGGCC CAATCAA 1041
Lettuce rbcL 2nd Half Reverse AATAATGAGC CAAGCTAGT- ATTTGCGGGG AATCCC---- ----CCA 1033
Consensus AANNNTNNNN CNNNNNANTG ANTTGCNNNN AANCNCGGCC CAATCNA
Conservation
100%
0%

2

L. sativa rbcL gene restriction sites using designed primers (forward strand)

```
                                       20                  40                  60                  80
Lettuce rbcL 2nd Half Forward  CTAAAATTGAAAGTATGGTCGTCCCTGTTGGGTGTACTATTAAACCTAAATTGGGGTTATCCGCTAAAAACTACGGCAGAGCTGTTTATGAA
                                    100                 120                 140                 160                 180
Lettuce rbcL 2nd Half Forward  TGTCTTCGTGGTGGCCTTGATTTTACTAAAGATGATGAGAACGTGAACTCCCAACCATTTATGCGTTGGAGAGACCGTTTCTTATTTTGTGC
                                        200                 220                 240                 260
Lettuce rbcL 2nd Half Forward  CGAAGCTATTTTTAAATCACAAGCTGAAACAGGTGAAATCAAAGGGCATTACTTGAATGCTACTGCGGGTACATGCGAAGAAATGATGAAAA
                                280                 300                 320                 340                 360
Lettuce rbcL 2nd Half Forward  GGGCTATATTTGCCAGAGAATTGGGAGTTCCTATCGTAATGCATGACTACCTAACAGGGGGATTCACTGCAAATACTAGCTTGGCTCATTAT
                                    380                 400                 420                 440                 460
Lettuce rbcL 2nd Half Forward  TGCCGAGATAATGGTCTACTTCTTCACATTCACCGCGAATCAAAAGGTCGAGATATGGTCTACTTCTTCCATCCACCGCGCAATGCATGCAG
                                         480                 500                 520                 540
Lettuce rbcL 2nd Half Forward  TTATTGATAGACAGAAGAATCATGGTATACACTTCCGTGTACTAGCTAAAGCGTTACGTATGTCTGGTGGAGATCATATTCATTCCGGTACC
                                    560                 580                 600                 620                 640
Lettuce rbcL 2nd Half Forward  GTAGTAGGTAAACTTGAAGGGGAAAGAGAAATCACTTTGGGCTTTGTTGATTTACTGCGTGATGATTTTATTGAAAAAGATAGAAGTCGCGG
                                        660                 680                 700                 720
Lettuce rbcL 2nd Half Forward  TATTTATTTCACCCAAGATTGGGTCTCTCTACCAGGTGTTCTGCCTGTAGCTTCGGGCGGTATTCACGTTTGGCATATGCCTGCTCTGACCG
                                740                 760                 780                 800                 820
                               BglII
Lettuce rbcL 2nd Half Forward  AGATCTTTGGAGATGATTCCGTACTACAGTTCGGTGGAGGAACTTTAGGGCACCCTTGGGGAAATGCACCCGGTGCCGTAGCTAATCGAGTA
                                    840                 860                 880                 900                 920
                                  XbaI
Lettuce rbcL 2nd Half Forward  GCTCTAGAAGCATGTGTACAGGCTCGTAATGAGGGACGCGATCTTGCTACTGAGGGTAATGAAATTATCCGTGAGGCTACCAAATGGAGTCC
                                         940                 960                 980                 1,000
Lettuce rbcL 2nd Half Forward  TGAACTAGCTGCTGCTTGTGAAGTATGGAAGGAGATTAAATTTGAGTTTCAGGCAATGGATACTTTGGATCAATAACTTTCGTTCTCTTAAT
                                    1,020               1,040
Lettuce rbcL 2nd Half Forward  TGAATTGCAATTAAACTCGGCCCAATCAA
```

L. sativa rbcL gene restriction sites using designed primers (reverse strand)

20 40 60 80

Lettuce rbcL 2nd Half Reverse **CGCCGGCATATGAGCAGCTAGTATTTGCAGTGATCCCCCTGTTAGGTAGTCATGCATTACGATAGGAACTCCCAATTCTCTGGCAAATATAG**

100 120 140 160 180

Lettuce rbcL 2nd Half Reverse **CCCTTTTCATCATTTCTTCGCATGTACCCGCAGTAGCATTCAAGTAATGCCCTTTGATTTCACCTGTTTCAGCTTGTGATTTAAAAATAGCT**

200 220 240 260

Lettuce rbcL 2nd Half Reverse **TCGGCACAAAATAAGAAACGGTCTCTCCAACGCATAAATGGTTGGGAGTTCACGTTCTCATCATCTTTAGTAAAATCAAGGCCACCACGAAG**

280 300 320 340 360

Lettuce rbcL 2nd Half Reverse **ACATTCATAAACAGCTCTACCGTAGTTTTTAGCGGATAACCCCAATTTAGGTTTAATAGTACATCCCAACAGGGGACGACCATACTTGTTCA**

380 400 420 440 460

BamHI

Lettuce rbcL 2nd Half Reverse **ATTTATCTCTCTCAACTTGGATCCAATGAGGCGATACGCGAACTAGTATTGATCCAAGTATCCATTGCCTGAACTCAAATTTAATCTCCTTC**

480 500 520 540

Lettuce rbcL 2nd Half Reverse **CATACTTCACAAGCAGCAGCTAGTTCAGGACTCCATTTGGTAGCCTCACGGATAATTTCATTACCCTCAGTAGCAAGATCGCGTCCCTCATT**

560 580 600 620 640

XbaI

Lettuce rbcL 2nd Half Reverse **ACGAGCCTGTACACATGCTTCTAGAGCTACTCGATTAGCTACGGCACCGGGTGCATTTCCCCAAGGGTGCCCTAAAGTTCCTCCACCGAACT**

660 680 700 720

BglII

Lettuce rbcL 2nd Half Reverse **GTAGTACGGAATCATCTCCAAAGATCTCGGTCAGAGCAGGCATATGCCAAACGTGAATACCGCCCGAAGCTACAGGCAGAACACCTGGTAGA**

740 760 780 800 820

Lettuce rbcL 2nd Half Reverse **GAGACCCAATCTTGGGTGAAATAAATACCGCGACTTCTATCTTTTTCAATAAAATCATCACGCAGTAAATCAACAAAGCCCAAAGTGATTTC**

840 860 880 900 920

Lettuce rbcL 2nd Half Reverse **TCTTTCCCCTTCAAGTTTACCTACTACGGTACCGGAATGAATATGATCTCCACCAGACATACGTAACGCTTTAGCTAGTACACGGAAGTGTA**

940 960 980 1,000

Lettuce rbcL 2nd Half Reverse **TACCATGATTCTTCTGTCTATCAATAACTGCATGCATTGCGCGGTGGATGTGAAGAAGTAGACCATTATCTCGGCAATAATGAGCCAAGCTA**

1,020

Lettuce rbcL 2nd Half Reverse **GTATTTGCGGGGAATCCCCCA**

Terence L. Stephens, MS

Email: Stephens.Terence@gmail.com

RESEARCH EXPERIENCE:

Eastern Washington University, Cheney WA Jan 2010–Mar 2010

Title: Research Assistant I position – Biology Dept.

- Grant funded research study using standard molecular techniques to analyze the virology of the Sigma-Rhabdo virus in Drosophila melanogaster. Data analysis of host/virus/vector interactions of the Sigma virus confirmed that genetic variation affects the level of infection in Drosophila melanogaster.

Eastern Washington University, Cheney WA Sept 2009–June 2012

Title: Graduate Research and Research Assistant I – Biology Dept.

- Examination of potential DNA barcodes for land plants. Two proposed barcodes, Ribulose-1,5-Bisphosphate Carboxylase/Oxygenase (RuBisCO) and maturase K, were amplified using the primers I designed and compared to a known reference library of DNA sequences via BLAST analysis. Barcode applicability were based on ease of amplification, discriminatory value, and sequence quality. Of the two proposed DNA barcodes, only RuBisCO satisfied all three parameters with little modification to enzymes and thermocycling conditions.

LABORATORY SKILLS AND TECHNIQUES

Molecular	Microbiological
– PCR optimization and trouble shooting – Primer design – Touchdown PCR – Recombinant DNA expression technologies – Cloning – Plasmid isolation and ligation – Sequencing reaction design and analysis	– Aseptic/sterile techniques – Experience with clinical pathogens – Culturing fastidious pathogens (bacterial and fungal) – API 20/biochemical tests – Characterization and identification of microbes – Bacterial assays (titers/gram test/ isolation) – Transformation <u>Bioinformatics</u> – BLAST analysis – Protein modeling – ClustalW – Primer3

ADDITIONAL SKILLS:

Communication and Computer skills:_

- MS Word, Excel, PowerPoint, NCBI, RefWorks, Endnote, Windows OS, Mac OS, Open Office, and multiple literature search databases
- Excellent record keeping (accurate and organized notebooks in school and laboratory positions
- Planning and executing research projects and data analysis (Masters and laboratory)
- Strong written, verbal, and presentation skills (thesis, abstracts, oral and poster presentations, teaching)

TEACHING EXPERIENCE:

Eastern Washington University, Cheney, Washington June 2010–Jun 2011

Graduate Teaching Assistant/Instructor

- Molecular Biology (3 quarters) – Assisted students in the laboratory with general molecular biology techniques eq. PCR, primer design, BLAST.
- Microbiology (2 quarters) – Assisted multiple laboratories of undergraduates in the pre-allied health and biotechnology field. Students learned basic principles and methods in microbiology.

EDUCATION

Graduate

Eastern Washington University, Cheney, WA Sept 2009–April 2013

M.S. Biological Sciences

Thesis: DNA Barcoding of Jatropha curcas

Awards

Eastern Washington University Symposium Award June 2010

Eastern Washington University Symposium Award June 2011

Undergraduate

Eastern Washington University, Cheney, WA Sept 2006–June 2009

B.S. Biological Sciences

Awards

Undergraduate Dean's List Dec 2008

Undergraduate Dean's List Mar 2009

Relevant Coursework (undergraduate and graduate)

Directly relevant coursework:	Supportive coursework:
– Molecular Biology with Lab (15 credits)	–Immunology (5 credits)
– Cellular Biology (5 credits)	–Epidemiology (5 credits)
– Genetics with Lab (4 credits)	–Virology (5 credits)
– Biological Investigation (3 credits)	–Chemistry with Lab (15 credits)
– Bioinformatics (3 credits)	

PROFESSIONAL DEVELOPMENT AND CONFERENCES:

EWU Spring Symposium, Cheney, WA (Oral Presenter) June 2010

EWU Quarterly Seminar, Cheney, Wa (Poster Presentation) Dec 2010

EWU Spring Symposium, Cheney, WA (Oral Presenter) June 2011